

**MODELO DE PREDICCIÓN DE EMERGENCIAS MÉDICAS UTILIZANDO
MÉTODOS DE APRENDIZAJE MÁQUINA ENFOCADOS EN EL ANÁLISIS
DE VARIABLES ESPACIALES Y TEMPORALES**

FRANS GUILLERMO TABOADA RIVERA

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

2024

**MODELO DE PREDICCIÓN DE EMERGENCIAS MÉDICAS UTILIZANDO
MÉTODOS DE APRENDIZAJE MÁQUINA ENFOCADOS EN EL ANÁLISIS
DE VARIABLES ESPACIALES Y TEMPORALES**

FRANS GUILLERMO TABOADA RIVERA

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

Director:

David Romo Bucheli, Doctor en ingeniería eléctrica

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

2024

AGRADECIMIENTOS

En esta sección deseo expresar mis más sinceros agradecimientos a quienes han sido fundamentales en el desarrollo de este proyecto de investigación. A mis padres, Victor y Maria, quienes han estado a mi lado en cada paso de este viaje, les debo un agradecimiento eterno. Su fe inquebrantable en mis capacidades y, sobre todo, su amor constante que me ha brindado las fuerzas para enfrentar cada uno de los desafíos que parecían imposibles. Cada sacrificio que han hecho para apoyar mis sueños es un recordatorio de su gran participación en este hito en mi vida. A mi hermano, Hans, por ser una inagotable fuente de apoyo durante toda mi carrera y vida.

A mi director, David Edmundo Romo Bucheli, cuya orientación y conocimiento me ha guiado en los momentos más cruciales de esta investigación. Su paciencia y dedicación con mi crecimiento académico y personal han sido invaluable y estoy sinceramente agradecido por todo su apoyo.

ÍNDICE GENERAL

	Pág
INTRODUCCIÓN	14
1. MARCO CONCEPTUAL	17
1.1. INSTAURACIÓN DE LOS SERVICIOS MÉDICOS PREHOSPITALARIOS	17
1.2. CLASIFICACIÓN DE EMERGENCIAS	19
1.3. CONTEXTO DE LOS SISTEMAS DE EMERGENCIAS EN COLOMBIA	21
2. TRABAJOS PREVIOS	23
2.1. MOTOR DE INTELIGENCIA ARTIFICIAL	23
2.2. TIPOS DE APRENDIZAJE AUTOMÁTICO	24
2.3. APRENDIZAJE AUTOMÁTICO DE REGRESIÓN	26
2.4. MODELOS ESPACIALES	27
2.4.1. Datos geoespaciales	28
2.4.2. Tipos de clases de entidades	28
2.4.3. Modelos de Predicción espacial	29
2.5. MODELOS TEMPORALES	30
2.5.1. Modelos de Predicción en series temporales	31
2.6. APLICACIÓN DE MODELOS ESPACIALES Y TEMPORALES	32
3. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA	36
4. OBJETIVOS	39
5. MATERIALES	40
5.1. ZONA DE ESTUDIO	40

5.1.1. Ubicación geográfica	40
5.1.2. Contexto Geoespacial	41
5.1.3. Datos Demográficos	41
5.1.4. Instituciones de interés	41
5.2. DESCRIPCIÓN DE FUENTES DE DATOS	43
5.2.1. Registro de emergencias médicas y accidentes de transito	43
5.2.2. Datos Geoespaciales	43
5.2.3. Datos demográficos	47
5.2.4. Datos climáticos	47
5.2.5. Datos temporales	51
5.2.6. Datos de tráfico y movilidad	52
5.3. PROCESAMIENTO DE DATOS	53
5.3.1. Limpieza de datos y estandarización	54
5.3.2. Geocodificación	54
5.3.3. Depuración espacial	55
5.4. CONTRUCCIÓN DEL DATASET	56
5.4.1. Conjunto de Datos del CRUE	57
5.4.2. Conjunto de Datos de la Dirección de Tránsito	57
5.5. ANÁLISIS EXPLORATORIO DE LAS EMERGENCIAS MEDICAS	58
6. METODOLOGÍA	64
6.1. REPRESENTACIÓN ESPACIAL	64
6.2. REPRESENTACIONES TEMPORALES	66
6.3. SIMULACIÓN DE TRÁFICO	69
6.4. FORMACIÓN DE VECTORES DE CARACTERÍSTICAS	72
6.5. MODELAMIENTO	73
7. MODELO DE APRENDIZAJE AUTOMÁTICO	76

8. EXPERIMENTOS Y RESULTADOS	80
8.1. CONFIGURACIÓN EXPERIMENTAL	80
8.1.1. Métricas de evaluación	80
8.1.2. Configuración de los vectores de características	81
8.2. EXPERIMENTOS	85
8.2.1. Experimentos con el Dataset CRUE	85
8.2.2. Experimentos con el Dataset DTB	86
8.3. RESULTADOS	87
8.3.1. Optimización de Hiperparámetros	87
8.3.2. Resultados de Experimentos	88
8.4. VISUALIZACIÓN DE RESULTADOS	90
9. CONCLUSIONES Y PERSPECTIVAS	101
BIBLIOGRAFÍA	103
ANEXOS	111

ÍNDICE DE FIGURAS

	Pág
Figura 1. Diagrama de funcionamiento SEM (2019). Las secciones de este sistema que son de interés para el proyecto son: 2. Determinación de requerimientos y coordinación de unidades. 3. Atención de la emergencia, recuperada y adaptada del Ministerio de Salud y Protección Social ¹	19
Figura 2. Tipos de aprendizaje automático. Creación propia. Creada y adaptada a partir de "Types of Machine Learning Algorithms" ²	24
Figura 3. Ruta de atención. Tomado de la página web del ministerio de salud y protección social de Colombia. estraído de la pagina del Ministerio de Salud y Protección Social ³	36
Figura 4. Área de estudio. Creación propia.	40
Figura 5. MGN de division politica del municipio de Bucaramnga y sus secciones urbanas. Creación propia con geopandas ⁴ y matplotlib. ⁵	46

¹ Ministerio de salud y proteccion social. *Sistema de emergencias médicas SEM*. Sistema de emergencias medicas-SEM. 29 de dic. de 2019. URL: <https://www.minsalud.gov.co/salud/PServicios/Paginas/Sistema-de-emergencias-medicas-SEM.aspx> (visitado 23-08-2023).

² Taiwo Ayodele. "Types of Machine Learning Algorithms". En: *New Advances in Machine Learning*. 1 de feb. de 2010. DOI: 10.5772/9385.

³ Ministerio de Salud y Protección Social. *Gestión integral de riesgo en salud*. Gestión integral de riesgo en salud. 19 de dic. de 2019. URL: <https://www.minsalud.gov.co/proteccionsocial/Paginas/gestion-integral-de-riesgo-en-salud.aspx> (visitado 29-08-2023).

⁴ Kelsey Jordahl et al. *geopandas/geopandas: v0.8.1*. Ver. v0.8.1. Jul. de 2020. DOI: 10.5281/zenodo.3946761.

⁵ J. D. Hunter. "Matplotlib: A 2D graphics environment". En: *Computing in Science & Engineering* 9.3 (2007), págs. 90-95. DOI: 10.1109/MCSE.2007.55.

Figura 6. Precipitación diaria en estaciones meteorológicas de Bucaramanga. Creación propia con pandas y matplotlib.	48
Figura 7. Histórico de variables climáticas en el área metropolitana de Bucaramanga. Creación propia.	50
Figura 8. Frecuencia de la dirección del viento en grados. Creación propia.	51
Figura 9. Días de importancia social. Creación propia con pandas y calmap.	52
Figura 10. Figura explicativa geocodificación. Creación propia.	55
Figura 11. Polígono Area metropolitana de Bucaramanga. Creación propia con OpenStreetMap.	56
Figura 12. Selección de dataset CRUE. Creación propia.	57
Figura 13. Selección de dataset DTB. Creación propia.	58
Figura 14. Emergencia médicas ocurridas en el área metropolitana de Bucaramanga. Creación por medio de la librería Matplotlib y análisis de tendencia con statsmodels.	60
Figura 15. Accidentes de tránsito en el área metropolitana de Bucaramanga. Creación por medio de la librería Matplotlib y análisis de tendencia con statsmodels.	60
Figura 16. Distribución de la emergencia médicas en el área metropolitana de Bucaramanga. Creación propia con seaborn y matplotlib.	63
Figura 17. Distribución de accidentes de tránsito en el área metropolitana de Bucaramanga. Creación propia con seaborn y matplotlib.	63
Figura 18. Comparación entre datos crudo, intersecciones originales y reducidas de Bucaramanga. Creación por medio de las librerías Matplotlib, OSMnx y Network.	65
Figura 19. Representación de codificación mensual cíclica, ejemplificación del mes de marzo. Creación propia.	68
Figura 20. Diagrama relacional. Creación propia con MySQL workbench.	71
Figura 21. Diagrama de flujo. Creación propia.	73
Figura 22. Un árbol de decisión que muestra la cantidad de accidentes el día siguiente en una intersección particular. las cifras debajo de las hojas muestran la cantidad.	77

Figura 23. Gráfico de dispersión que compara las observaciones reales y las observaciones predichas por el modelo CRUE. Ambos conjuntos de datos están normalizados utilizando la escala min-max. Los resultados obtenidos se reflejan en las bajas métricas de Error Cuadrático Medio (MSE) de 0.006 y un coeficiente de determinación de 0.334. Estos valores indican una precisión notable y un ajuste efectivo del modelo para explicar la variabilidad en los datos durante el período analizado	91
Figura 24. El gráfico de las 5 mejores características, se construyó mediante el algoritmo de Bosques Aleatorios utilizando el conjunto de datos CRUE. Se calculó la importancia de cada característica evaluando su contribución a la capacidad predictiva del modelo. En la gráfica, cada barra horizontal representa una característica, y su longitud indica la magnitud de la importancia. Las características se han etiquetado en el eje vertical. Cuanto más alta sea la barra, mayor será la importancia de la característica correspondiente. Las características destacadas son la suma total de los últimos 14 días y los últimos 5 días	92
Figura 25. Serie temporal de entrenamiento, test y predicciones. conjunto de datos CRUE .	93
Figura 26. Serie temporal de datos acumulativos de test y predicciones, conjunto de datos CRUE	94
Figura 27. Mapas de calor en la zona de estudio de los datos de prueba y predicciones, del conjunto de datos CRUE	95
Figura 28. 5 mejores características en el modelo de la DTB(2017-2019)	96
Figura 29. 5 mejores características en el modelo de la DTB(2021-2022)	96
Figura 30. Serie temporal de entrenamiento, testeo y predicciones, conjunto de datos DTB(2017-2019)	97
Figura 31. Serie temporal de entrenamiento, testeo y predicciones, conjunto de datos DTB(2021-2022)	97
Figura 32. Serie temporal de testeo y predicciones, conjunto de datos CRUE	98
Figura 33. Serie temporal de testeo y predicciones, conjunto de datos DTB(2017-2019) . .	98

Figura 34.Mapas de calor de datos de prueba y predicciones, conjunto de datos DTB(2017-2019) 99

Figura 35.Mapas de calor de datos de prueba y predicciones, conjunto de datos DTB(2021-2022) 100

ÍNDICE DE CUADROS

	Pág
Tabla 1. Categorías y criterios de triaje (2012). Recuperado de ⁶	20
Tabla 2. Magnitud del evento. Recuperado de ⁷	20
Tabla 3. Codificación de Experimentos CRUE	85
Tabla 4. Codificación de Experimentos DTB (2017-2019)	86
Tabla 5. Codificación de Experimentos DTB (2021-2022)	86
Tabla 6. Resultados de la optimización de hiperparámetros para el modelo CRUE.	87
Tabla 7. Resultados de la optimización de hiperparámetros para el modelo STB (2017-2019).	88
Tabla 8. Resultados de la optimización de hiperparámetros para el modelo DTB (2021- 2022).	88
Tabla 9. Resultados de los experimentos para el modelo CRUE.	89
Tabla 10. Resultados de los experimentos para el modelo DTB (2017-2019).	89
Tabla 11. Resultados de los experimentos para el modelo DTB (2021-2022).	90

⁶ Universidad de Antioquia y Ministerio de Salud y Protección Social. *Guías Básicas de Atención Médica Prehospitalaria*. 2012.

⁷ Universidad de Antioquia y Ministerio de Salud y Protección Social. *Guías Básicas de Atención Médica Prehospitalaria*. 2012.

RESUMEN

TÍTULO: MODELO DE PREDICCIÓN DE EMERGENCIAS MÉDICAS UTILIZANDO MÉTODOS DE APRENDIZAJE MÁQUINA ENFOCADOS EN EL ANÁLISIS DE VARIABLES ESPACIALES Y TEMPORALES. *

AUTOR: FRANS GUILLERMO TABOADA RIVERA **

PALABRAS CLAVE: Servicios prehospitalarios, Machine learning, Salud, Variables espaciales y temporales.

DESCRIPCIÓN: Este trabajo de investigación se enfoca en el desarrollo de un modelo de predicción de emergencias médicas en el área metropolitana de Bucaramanga, Colombia, haciendo uso de métodos de aprendizaje máquina y análisis de variables espaciales y temporales. La metodología implementada se basa en el concepto de ventana deslizante para el análisis y modelado de series temporales, y se realizaron experimentos exhaustivos para ajustar la configuración final del modelo.

Los resultados obtenidos respaldan la viabilidad de establecer relaciones entre variables temporales y geográficas, especialmente en el contexto crítico de la atención a emergencias médicas. Los datos utilizados provienen principalmente del Centro Regulador de Urgencias y Emergencias del departamento de Santander y la Dirección de Tránsito de Bucaramanga. Se aplicaron procesos como la geocodificación y el relacionamiento espacial para generar vectores de características que incorporaran dependencia espacial y temporal.

Además de los datos principales, se recopiló información demográfica, climatológica y de contexto social para evaluar su contribución al modelo. La investigación abarcó más de 40 experimentos, involucrando diversas configuraciones de ventanas deslizantes, pruebas con distintos métodos de normalización y un análisis detallado de las variables resultantes de la ingeniería de características.

Este estudio no solo demuestra la capacidad del modelo para prever emergencias médicas, sino que también subraya la importancia de considerar factores temporales y espaciales en la toma de decisiones en el ámbito de la respuesta inmediata. La combinación de técnicas de aprendizaje máquina, análisis geoespacial y la inclusión de variables complementarias fortalece la aplicabilidad y relevancia de este enfoque en el contexto de la atención a emergencias en entornos urbanos.

* Trabajo de investigación

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: David Romo Bucheli

ABSTRACT

TITLE: MEDICAL EMERGENCY PREDICTION MODEL USING MACHINE LEARNING METHODS FOCUSED ON THE ANALYSIS OF SPATIAL AND TEMPORAL VARIABLES. *

AUTHOR: FRANS GUILLERMO TABOADA RIVERA **

KEYWORDS: Pre-hospital services, Machine learning, Health, spatial and temporal variables

DESCRIPTION: This research work focuses on the development of a prediction model for medical emergencies in the metropolitan area of Bucaramanga, Colombia, using machine learning methods and analysis of spatial and temporal variables. The implemented methodology is based on the sliding window concept for time series analysis and modeling, and extensive experiments were performed to fine-tune the final model configuration.

The results obtained support the viability of establishing relationships between temporal and geographical variables, especially in the critical context of medical emergency care. The data used comes mainly from the Emergency and Urgency Regulator Center of the Department of Santander and the Transit Authority of Bucaramanga. Processes such as geocoding and spatial relationships were applied to generate feature vectors that incorporated spatial and temporal dependence.

In addition to the main data, demographic, climatological, and social context information was collected to evaluate their contribution to the model. The research spanned more than 40 experiments, involving various sliding window configurations, tests with different normalization methods, and a detailed analysis of the variables resulting from feature engineering.

This study not only demonstrates the model's ability to predict medical emergencies but also highlights the importance of considering temporal and spatial factors in decision-making in the immediate response area. The combination of machine learning techniques, geospatial analysis, and the inclusion of complementary variables strengthens the applicability and relevance of this approach in the context of emergency response in urban environments.

* Research work

** Faculty of Physical-Mechanical Engineering. School of Systems and Computer Engineering. Advisor: David Romo Bucheli

INTRODUCCIÓN

En Colombia, el sistema de emergencias médicas (SEM) hace parte fundamental de la atención integral a los ciudadanos. Este sistema busca la atención y respuesta oportuna a situaciones fortuitas que requieran de una asistencia médica inmediata. Algunas de estas son: afecciones por enfermedades, desastres naturales, accidentes de tránsito, traumatismos, emergencias cardiovasculares, entre otras¹. Para subrayar la importancia de contar con un SEM eficaz, podemos mencionar que la ocurrencia de accidentes de tránsito en Colombia entre enero de 2021 y junio de 2022 asciende a más de 10.000 casos ².

Un par de aspectos cruciales del SEM son la notificación y gestión de recursos (ambulancias) para el traslado a centros hospitalarios o de atención inmediata¹. Estos factores afectan de forma determinante el cuadro médico de los individuos que acceden a este servicio. No obstante, la atención prehospitalaria es multifactorial y no podemos atribuir todo el resultado del servicio al SEM, pero sí podemos identificar las problemáticas que afrontan y tratarlas con diversos enfoques.

La distribución desigual de los recursos de atención prehospitalaria presenta una problemática crítica dentro de las urbes, afectando directamente a la calidad de este servicio y su capacidad de respuesta. Según la Organización Mundial de la Salud (OMS), cada año se atribuyen entre 5,7 a 8,4 millones de muertes a la deficiente calidad de la atención en los servicios de salud en países de ingresos bajos y medianos como Colombia. Estas cifras podrían representar hasta el

¹ Ministerio de salud y protección social. *Resolución Número 926 de 2017*. 30 de mar. de 2017.

² Agencia Nacional de Seguridad Vial. *Histórico víctimas | ANSV*. Histórico víctimas | ANSV. URL: <https://ansv.gov.co/es/observatorio/estad%C3%ADsticas/historico-victimas> (visitado 22-08-2023).

15% de las muertes totales en estos territorios³.

La importancia de prever emergencias hospitalarias se vuelve evidente en este contexto, ya que la capacidad de anticipar y gestionar eficazmente la demanda de atención médica puede marcar la diferencia en la vida de las personas. La predicción de emergencias hospitalarias ofrece la posibilidad de mejorar la distribución de recursos, permitiendo una respuesta más eficiente ante situaciones críticas. Además, este enfoque podría contribuir a la planificación de bases de acción inmediata en áreas propensas a emergencias, mejorando la capacidad de respuesta ante eventos inesperados.

El objetivo de este trabajo de grado es proporcionar una representación espacial con marcadores temporales adaptada a las necesidades del SEM. De esta manera, la ocurrencia y el comportamiento de las emergencias médicas podrá ser representada de forma clara durante un periodo de tiempo específico. Este marco de representación permitirá además el desarrollo de modelos de predicción de emergencias, los cuales facilitarían la identificación de patrones en la ocurrencia de estas, y así conseguir información importante para la toma de decisiones, buscando obtener una administración eficaz de los recursos.

Algunas de las ventajas que ofrece este proyecto de investigación frente a productos comerciales como ArcGis⁴ y QGIS⁵ son: la posibilidad del uso del modelo como base pre-entrenada para futuros abordajes a problemáticas en esta zona geográfica específica, esto debido a que el nivel de personalización al que es acotado el modelo podría brindar un resultado más preciso y adap-

³ Organización Mundial de la Salud. *Servicios sanitarios de calidad*. Servicios sanitarios de calidad. 11 de ago. de 2020. URL: <https://www.who.int/es/news-room/fact-sheets/detail/quality-health-services> (visitado 22-08-2023).

⁴ Esri. *ArcGIS*. Ver. 10.8.2. 9 de dic. de 2021.

⁵ Equipo desarrollador QGIS. *QGIS*. Ver. 3.32.2. 23 de jun. de 2023.

tado a las particularidades de la zona. El lenguaje de programación utilizado para el desarrollo (python), actualmente está adaptado de manera nativa para su uso en arquitecturas web permitiendo así su correcta manipulación para el despliegue de servicios de predicción embebidos en los sistemas de información particulares. A su vez, este tipo de lenguajes interpretados presentan una ventaja en el uso de recursos computacionales, ya que elimina la necesidad de herramientas más pesadas y complejas.

1. MARCO CONCEPTUAL

En esta sección se exponen los principales conceptos asociados a los sistemas de emergencias médicas prehospitales, junto con una descripción de las principales variables, mecanismos, marco regulatorio, etc. Esto para la comprensión general del proyecto y su contextualización.

1.1. INSTAURACIÓN DE LOS SERVICIOS MÉDICOS PREHOSPITALARIOS

El sistema de salud colombiano está compuesto por diversas entidades que son financiadas por recursos públicos y privados. En Colombia, el sector que conglomeró todos estos servicios es el Sistema General de Seguridad Social en Salud (SGSSS). Para acceder a este sistema se debe hacer por medio de Entidades Promotoras de Salud (EPS), los usuarios se afilian a estas instituciones prestadoras de servicios (IPS) por medio de regímenes, los cuales se dividen en régimen contributivo (RC) y régimen subsidiado (RS), además, la afiliación es obligatoria para todos los colombianos. Las EPS tienen como proveedores de servicios a las IPS, que pueden estar o no integradas a las EPS pero que en cualquier caso son contratadas por estas. Los servicios que prestan las IPS deben obedecer al Plan Obligatorio de Salud (POS) o bien el Plan Obligatorio de Salud-subsidiado (POS-S) para los afiliados al RS. Este plan contiene las atenciones básicas y primordiales que se deben prestar a cualquier ciudadano colombiano, incluyendo el servicio de atención prehospitalaria.

Una parte importante del SGSSS, corresponde a los sistemas de emergencia en atención prehospitalaria a nivel local que interactúan con las instituciones prestadoras de servicios de salud, que están a cargo del despliegue de los recursos para atender las emergencias. En la resolución 926 de 2017 se reglamenta el Sistema de Emergencias Médicas (SEM), el cual agrupa diversas entidades de direccionamiento, coordinación no asistencial y operadores asistenciales de cada región. Esta resolución contiene una serie de disposiciones y normas que se aplican en la dirección de las diferentes organizaciones involucradas en la prestación del servicio¹. Algunas de las

funciones del SEM son:

1. Garantizar la articulación del Centro Regulador de Urgencias y Emergencias CRUE con el Número Único de Seguridad y Emergencias (NUSE) o aquel que cumpla sus funciones.
2. Articular a los integrantes del SEM ante situaciones de emergencia o desastre en el marco del Sistema Nacional de Gestión del Riesgo de Desastres.
3. Promover programas de educación a la comunidad como primeros respondientes comunitarios ante emergencias; para tal fin, la entidad territorial podrá establecer alianzas con las demás entidades del Sistema Nacional de Gestión del Riesgo de Desastres o con la empresa privada.
4. Reportar al Ministerio de Salud y Protección Social la información que se requiera a través de los mecanismos que para el efecto se defina.
5. Responder a las necesidades de atención en salud de la población afectada por situaciones de urgencia, emergencia o desastre, incluida la atención prehospitalaria, transporte básico y medicalizado de pacientes, atención de urgencias y hospitalaria.
6. Promover la formación y capacitación del talento humano para cubrir las necesidades del SEM.
7. Apoyar los procesos de vigilancia epidemiológica.

Como parte fundamental para la prestación del servicio de emergencias se aclara que el usuario puede acceder a estos servicios sin previo aviso.

En la figura1, se representa de manera general el funcionamiento del SEM y los diferentes mecanismos necesarios para atender las emergencias prehospitalarias. El primer mecanismo consiste en la recepción de los llamados de emergencia en los distintos canales de comunicación. Posteriormente, como segundo mecanismo se determina las necesidades, se identifican las ambulancias más cercanas y se asigna la atención de la emergencia. A continuación se realizan todos los procedimientos que brinda la ambulancia en el momento de la atención de la emergencia. El cuarto mecanismo le corresponde a la atención en el centro de salud específico que se asignó dependiendo de la gravedad de la emergencia. En ese

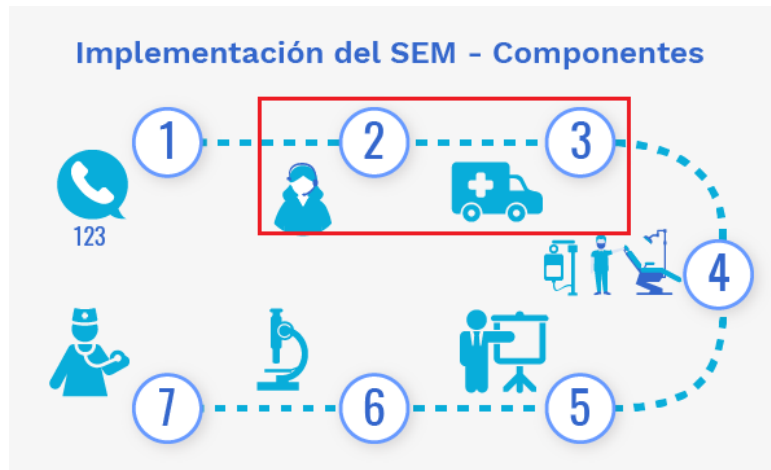


Figura 1. Diagrama de funcionamiento SEM (2019). Las secciones de este sistema que son de interés para el proyecto son: 2. Determinación de requerimientos y coordinación de unidades. 3. Atención de la emergencia, recuperada y adaptada del Ministerio de Salud y Protección Social ⁶.

mismo centro de salud se realiza el diagnóstico y tratamiento. Mecanismos adicionales están asociados a los programas educacionales y los procesos de vigilancia orientados a reducir la ocurrencia de las emergencias.

1.2. CLASIFICACIÓN DE EMERGENCIAS

Las acciones que permiten clasificar una emergencia se denominan triaje, que se define como “el proceso de categorización de lesionados basado en la urgencia de sus lesiones y la posibilidad de supervivencia”⁷. Dicho proceso implica la evaluación de las lesiones, la asignación de prioridad, la identificación de factores de riesgo, entre otros.

La clasificación de lesionados se ha estandarizado de la siguiente forma:

Al finalizar la identificación de los lesionados mediante la tabla 1, todo el evento se debe clasificar según su grado de magnitud, debido a que las instituciones hospitalarias deben adoptar medidas para el manejo de la situación.

⁷ Universidad de Antioquia y Ministerio de Salud y Protección Social. *Guías Básicas de Atención Médica Prehospitalaria*. 2012.

Prioridad	Descripción
I	Las lesiones que posee son de cuidados intermedios y requieren de atención inmediata.
II	Las lesiones que posee son diferibles y su atención médica puede dar espera.
III	Las lesiones que poseen son de alta gravedad y existen pocas posibilidades de supervivencia.
IV	Las lesiones que posee son leves y su atención puede dejarse para el final.
V	Se utiliza cuando la persona falleció.

Tabla 1. Categorías y criterios de triaje (2012). Recuperado de ⁸

Situación 1	Situación 2	Situación 3	Tipo de lesiones	Categoría
X			No lesionados	Verde
X	X		Fracturas simples	Verde
X	X	X	Fractura múltiples y severas	Amarilla
X	X	X	Dificultad respiratoria	Roja
X	X	X	Hemorragias en la caja torácica	Roja
X	X		Compromiso severo de órganos	Negra
X			Trauma encefalocraneano	Negra
			Muerte	Blanca

Tabla 2. Magnitud del evento. Recuperado de ⁹

La clasificación representada en la tabla 2 expone tres diversas magnitudes del evento, las cuales son cruciales para la toma de decisiones en el sistema de emergencias.

Evento adverso compensado (Situación 1) Se refiere a situaciones donde los lesionados pese a su estado pueden ser atendidos de manera normal y la cadena de socorro no es necesaria, un ejemplo son los accidentes de tránsito o emergencias médicas generales.

Evento adverso descompensado (Situación 2) En el evento adverso descompensado, la magnitud aumenta y es necesaria la utilización de la cadena de socorro debido a que múltiples lesionados requieren de atención inmediata para su supervivencia, pero con la activación de planes hospitalarios de emergencia se puede garantizar el bienestar de los lesionados.

Desastre con multitud de lesionados (Situación 3) En el instante en el que un evento de magnitud mayor es notificado, solo los pacientes con lesiones mayores serán atendidos. Todos los recursos disponibles son utilizados en los lesionados que los requieran, a su vez el sistema hospitalario se preparara para activar protocolos de emergencia, los pacientes que no puedan ser atendidos en centrales

médicas serán tratados a nivel prehospitalario.

1.3. CONTEXTO DE LOS SISTEMAS DE EMERGENCIAS EN COLOMBIA

Para la comprensión de la problemática se debe descomponer la estructura y funcionamiento de los sistemas de emergencias en Colombia. El objetivo principal del sistema de emergencias es responder los eventos clasificados como emergencias de diversas índoles como accidentes de tránsito, enfermedades generales, traumatismos, paros cardiorrespiratorios o cualquier otra situación que amerite profesionales de salud. Este sistema se estructura por tres órganos principales los cuales son: direccionamiento, coordinación no asistencial y operadores asistenciales.

La **sección de direccionamiento** es compuesta a su vez por el Ministerio de Salud y Protección Social de estado y el Comité Nacional de Urgencias, los cuales se encargan de gestionar y coordinar.

La **coordinación no asistencial** está encabezada por los entes territoriales a través del Centro de Regulación de Urgencias, Emergencias y Desastres, los cuales articulan la cadena de emergencias en diversos ámbitos.

Los **operadores asistenciales** son entidades prestadoras de servicios de salud que, en este caso, atienden situaciones de emergencia. Lo anterior debe cumplirse de manera oportuna, eficiente y con alta calidad, enfocado a enfermedades generales, accidentes de tránsito, traumatismos o paros cardiacos.

En este proyecto se tendrá en cuenta principalmente una de las funciones que desempeñan los operadores asistenciales, la cual es brindar atención oportuna ante urgencias, emergencias o desastres bajo los servicios de atención prehospitalaria, traslado básico y/o medicalizado de los pacientes¹.

Interoperabilidad en los sistemas de salud La interoperabilidad se refiere a la capacidad de diversos productos tecnológicos tales como sistemas de información, aplicaciones de software, redes de comunicación, entre otros, de intercambiar eficientemente información de alto valor para el negocio.

El gran reto en el sector salud que se ha presentado en los últimos años es compartir y acceder a los datos de forma segura, esto debido a la confidencialidad que poseen junto con los altos requerimientos de privacidad y seguridad. La limitación en el acceso de dicha información podría afectar la prestación del servicio, ya que, la comprensión incompleta de las necesidades de un individuo o individuos podría ocasionar resultados insuficientes y costos elevados.

Debido al envejecimiento poblacional, y el aumento de la vida media, el intercambio de datos cumplirá un papel importante en la atención médica efectiva. Según el informe del Observatorio Nacional de Salud de Colombia para el 2011 un 71 % del total de la población requirió de una consulta médica al menos una vez en el año¹⁰

¹⁰ Ministerio de Salud y Protección Social, Instituto Nacional de Salud y Observatorio Nacional de Salud. *Aspectos relacionados con la frecuencia de uso de los servicios de salud, mortalidad y discapacidad en Colombia*. 2011.

2. TRABAJOS PREVIOS

El aprendizaje automático es un componente de la inteligencia artificial, el cual se centra en sistemas de aprendizaje o mejoramiento en el rendimiento en función de los datos de consumo. El aprendizaje de máquina se considera inteligencia artificial pero no todos los modelos de inteligencia artificial son aprendizaje de máquina.

En la actualidad el aprendizaje automático se hace presente en muchos aspectos cotidianos tales como la interacción con bancos, compras en línea, redes sociales o servicios de streaming.

2.1. MOTOR DE INTELIGENCIA ARTIFICIAL

Un motor de inteligencia artificial es una herramienta que ayuda a construir un sistema de inteligencia artificial. Estas herramientas ayudan a realizar tareas que son repetitivas y, a menudo, difíciles de realizar por parte de un ser humano. Hay una variedad de herramientas que están presentes actualmente en el mercado.

Estos motores se componen de diversos módulos según las necesidades del servicio, pero comparten un objetivo general el cual es facilitar tareas, junto con proporcionar un valor agregado a las empresas que los utilizan en sus funciones. De la misma forma, podrían estar enfocados en aumento de ventas o uso del servicio, reducciones de costos, mejora en el cumplimiento del servicio, mejoramiento de la seguridad, entre otros sectores decisivos del funcionamiento de la organización.

2.2. TIPOS DE APRENDIZAJE AUTOMÁTICO

Los algoritmos que permiten el desarrollo y funcionamiento del aprendizaje automático son divididos de la siguiente manera:

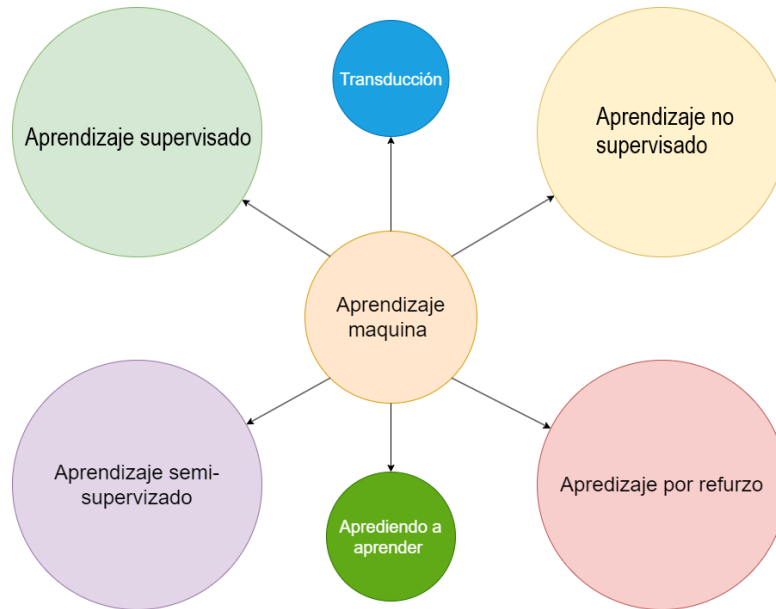


Figura 2. Tipos de aprendizaje automático. Creación propia. Creada y adaptada a partir de "Types of Machine Learning Algorithms"¹¹

Las diferentes formas de abordar problemáticas son muy diversas y con diferente complejidad, estas son representadas en la figura 2.

El aprendizaje automático es una de las ramas de la inteligencia artificial que ha tenido un crecimiento vertiginoso en los últimos años, esto debido a las novedosas aplicaciones de gran alcance que se han producido y han cautivado al público general. Este campo permite desarrollar modelos que pueden aprender a realizar predicciones y clasificaciones para el soporte de toma de decisiones basada en datos. El proceso generalizado del aprendizaje automático toma cierto tipo de entradas para determinar patrones y relaciones dentro de las mismas, por lo tanto, podrá replicar lo identificado en futuras entradas dando así un modelo relativamente preciso entrenado para tareas particulares.

Las principales metodologías según Taiwo Ayodele¹² son:

1. **En el aprendizaje supervisado** los datos utilizados para el entrenamiento que alimenta a los algoritmos incluyen etiquetas objetivo. Típicamente estos algoritmos son utilizados para clasificación de datos o predicción de resultados con precisión. A medida que los datos de entrada se introducen en el modelo, este ajusta sus pesos hasta que el modelo obtenga un menor error en la métrica de desempeño sobre el conjunto de datos de entrenamiento.
2. **El aprendizaje no supervisado** utiliza algoritmos de aprendizaje automático para analizar y agrupar conjuntos de datos no etiquetados. Estos algoritmos descubren patrones ocultos o agrupaciones de datos sin necesidad de intervención humana. Su capacidad para descubrir similitudes y diferencias en la información lo convierte en la solución ideal para el análisis exploratorio de datos.
3. **El aprendizaje reforzado** consiste en el desarrollo de algoritmos y modelos que permitan a un agente observar el entorno, seleccionar y realizar acciones para obtener recompensas o sanciones. El agente aprende por sí mismo realizando una maximización de la recompensa.
4. **El aprendizaje semisupervisado** tiene como objetivo el obtener información relevante a partir de datos no etiquetados. Los modelos son entrenados con datos etiquetados y un conjunto más grande de datos no etiquetados. Los datos etiquetados proporcionan una guía al modelo que se encarga de reconocer patrones y relaciones.
5. **Transducción** Posee ciertos rasgos compartidos con el aprendizaje supervisado, sin embargo, su propósito no es crear explícitamente características, sino simplemente intentar predecir futuras características basándose en entradas y nuevas variables. Este tipo de aprendizaje también es mencionado como aprendizaje supervisado dinámico.

¹² Taiwo Ayodele. "Types of Machine Learning Algorithms". En: *New Advances in Machine Learning*. 1 de feb. de 2010. DOI: 10.5772/9385.

6. **Aprendiendo a aprender** se refiere a la capacidad de un algoritmo para ajustar sus suposiciones sobre patrones de datos a medida que adquiere experiencia con ejemplos. A medida que encuentran más datos y situaciones diferentes, los algoritmos cambian gradualmente su conocimiento previo, lo que les permite comprender mejor la complejidad de los datos del mundo real y tomar decisiones más precisas en nuevos entornos.

2.3. APRENDIZAJE AUTOMÁTICO DE REGRESIÓN

La regresión es una técnica estadística utilizada para analizar y comparar la relación entre una variable dependiente y una o más variables independientes. El propósito del análisis de regresión es predecir el valor de la variable dependiente sobre el valor de las variables independientes.

La regresión se utiliza a menudo en varios campos, incluidos la economía, las finanzas, las ciencias sociales, la ingeniería y las ciencias naturales.

Este estudio se centra en la exploración de algoritmos de regresión, por lo tanto se definen algunos de los más utilizados a continuación.

1. **La regresión lineal** es un tipo de análisis predictivo el cual utiliza el comportamiento de una variable la cual está en función de otra, los nombres respectivos son variable dependiente e independiente.
2. **La regresión polinomial** toma la premisa de la regresión lineal pero en esta existen múltiples variables.
3. **La regresión de árboles de decisión** funciona dividiendo recursivamente el espacio de entrada en regiones más pequeñas y luego haciendo predicciones basadas en el valor promedio o mediano de la variable de destino en cada región. La principal ventaja de la regresión del árbol de decisión es su interpretabilidad. La estructura del árbol de decisiones nos permite visualizar y comprender fácilmente el razonamiento detrás de las predicciones, además, los árboles de decisión pueden manejar datos numéricos y categóricos, y pueden manejar valores faltantes.

4. **La regresión por bosques aleatorios** es un método de conjunto que combina múltiples árboles de decisión para hacer predicciones. La principal ventaja de la regresión por bosques aleatorios es que reduce el sobreajuste al combinar múltiples árboles de decisión. También puede manejar datos faltantes y puede manejar datos numéricos y categóricos. La regresión por bosques aleatorios es fácil de usar y no requiere una extensa ingeniería de funciones.
5. **Las máquinas de vectores de soporte de regresión (SVM)** se utilizan para predecir valores numéricos continuos. La base utilizada para el funcionamiento de las SVM es encontrar un hiperplano que separe los puntos de datos de tal manera que se maximice la distancia entre el hiperplano y los puntos de datos. Este hiperplano luego se usa para hacer predicciones para nuevos puntos de datos.

El algoritmo SVM de regresión funciona asignando los puntos de datos de entrada a un espacio de alta dimensión mediante una función kernel, lo que permite que el algoritmo encuentre un hiperplano que separe los puntos de datos en este espacio. Luego, el algoritmo resuelve un problema de optimización para encontrar el hiperplano que maximiza el margen, o la distancia entre el hiperplano y los puntos de datos más cercanos.

2.4. MODELOS ESPACIALES

Los modelos espaciales presentan una estructura con alta flexibilidad y gran simplicidad, lo cual permite abordar una amplia gama de problemas. Estos modelos permiten el manejo de diversos tipos de datos, ya sean continuos o discretos, así como datos de área o punto en el espacio. Por otro lado, su adaptación a disposiciones regulares o irregulares es una ventaja sustancial en la adaptación a diversos casos de estudio. Estos modelos se utilizan para analizar datos que presentan autocorrelación espacial, es decir, datos en los que los valores de una variable en una ubicación geográfica están relacionados con los valores de la misma variable en ubicaciones geográficas cercanas. Los modelos espaciales pueden ayudar a comprender mejor las relaciones entre variables en un contexto geográfico y a predecir valores

de variables en ubicaciones geográficas no observadas.¹³

2.4.1. Datos geoespaciales Como se menciona en el capítulo 10 página 171 de Robert Layton y Paul A. Watters¹⁴, los datos geoespaciales son datos referentes a objetos, eventos, fenómenos que tienen lugar en una ubicación en la superficie de la tierra. La ubicación puede llegar a ser estática o dinámica. Los datos geoespaciales poseen información de su ubicación en el espacio y a menudo también información temporal. Dada las características de estos datos, son el pilar para la presente investigación.

EPSG son siglas en inglés de “European Petroleum Survey Group”, que se traduce al español como “Grupo de Estudio de Petróleo Europeo”. Este término es más conocido por su relación con el estándar de identificación de sistemas de referencias espacial y proyecciones cartográficas. El EPSG Geodetic Parameter Dataset, es un registro en el que confluyen datos geodésicos, sistemas de referencia espacial, elipsoides terrestres, transformaciones de coordenadas y unidades de medida relacionadas. Los anteriores registros son utilizados en la actualidad para referirse de manera única a sistemas de referencia espacial en aplicaciones relacionadas con la cartografía, la geodesia y los sistemas de información geográfica (SIG). Cada uno de los conjuntos EPSG están estrechamente relacionados con parámetros descriptivos asociados al mapeo de la superficie terrestre en una representación específica. Esto es esencial para garantizar la precisión y la coherencia en la representación de datos geoespaciales en diferentes aplicaciones y sistemas. En el presente estudio se trabajan representaciones puntuales y de área, que gracias a EPSG se define su correcta transformación en el momento de su manipulación y tratamiento.

2.4.2. Tipos de clases de entidades

¹³ Coro Chasco. “Geodaspace: A Resource for Teaching Spatial Regression Models”. En: *Rect@* 4 (2013). Num Pages: 26 Place: Valencia, Spain Publisher: Ramón Sala Garrido, págs. 119-144.

¹⁴ Robert Layton y Paul A. Watters. *Automating Open Source Intelligence: Algorithms for OSINT*. San Diego, UNITED STATES: Elsevier Science & Technology Books, 2015.

Píxeles En el contexto de los datos raster, los píxeles son los elementos más pequeños de una imagen digital georreferenciada. Cada píxel contiene información sobre el valor de un atributo en una ubicación específica.

Puntos Los puntos en un SIG representan ubicaciones puntuales en la superficie de la Tierra, como las coordenadas de una ciudad, una tienda o cualquier otro punto de interés. Cada punto tiene coordenadas geográficas (latitud y longitud) que lo definen de manera única.

Líneas Las líneas son objetos lineales en un SIG que pueden representar características como carreteras, ríos, límites políticos, entre otros. Están definidas por una serie de coordenadas de inicio y fin, y pueden tener atributos adicionales como el tipo de vía o el nombre del río.

Polígonos Los polígonos son áreas definidas por una serie de vértices conectados en un orden específico. Representan regiones en un mapa y se utilizan para delinear características como países, estados, cuerpos de agua y zonas geográficas especiales.

2.4.3. Modelos de Predicción espacial

El Kriging es un método de interpolación que estima el valor de un punto no muestreado como una combinación ponderada de los valores de puntos circundantes, teniendo en cuenta la estructura de correlación espacial. Este método es ampliamente utilizado en la predicción de valores en ubicaciones no muestreadas y ha demostrado su eficacia en diversos campos geospaciales¹⁵.

La regresión espacial combina la regresión clásica con la consideración de la estructura espacial de los datos. Modelos como el SAR (Regresión Espacial Autorregresiva) y el CAR (Regresión Espacial Condicionada) son ejemplos de enfoques que incorporan la autocorrelación espacial en la relación entre las variables¹⁶.

¹⁵ Bostan P. “Basic Kriging Methods in Geostatistics”. En: *Yuzuncu Yil University Journal of Agricultural Sciences* 27 (31 de mar. de 2017).

¹⁶ Raymond J.G.M. Florax y Peter Nijkamp. “Misspecification in Linear Spatial Regression Models”. En: *Tinbergen Institute Discussion Papers* 2003-081/3 (3 de ago. de 2003).

Modelos de Máquinas de Soporte Vectorial (SVM) son un tipo de modelo de aprendizaje automático que también se ha aplicado con éxito en problemas de predicción espacial. Estos modelos buscan encontrar un hiperplano óptimo que separe las clases en el espacio de características, y su extensión a problemas espaciales ha demostrado ser efectiva¹⁷.

Bosques aleatorios destacan por su capacidad para manejar no linealidades, adaptarse a conjuntos de datos complejos y capturar interacciones espaciales. Al combinar múltiples árboles de decisión, los bosques aleatorios generan predicciones robustas, siendo una elección popular para problemas de predicción espacial¹⁸.

Redes Neuronales para Datos Espaciales Las redes neuronales, especialmente las redes neuronales convolucionales (CNN) en el ámbito de la visión por computadora, se han utilizado cada vez más en la predicción espacial. Estos modelos son capaces de capturar patrones complejos en datos geoespaciales y han mostrado buen rendimiento en diversas aplicaciones¹⁹.

2.5. MODELOS TEMPORALES

En esta sección se describirán algunos de los modelos más utilizados en la predicción de series temporales. Antes de abordar estos modelos es pertinente una explicación simple al momento de referirse a una serie temporal.

¹⁷ Penzuti Pacheco P. Narrandes S. Wang Y. Huang S. Cai N. y Xu W. “Applications of Support Vector Machine (SVM) Learning in Cancer Genomics”. En: *Cancer Genomics - Proteomics* 15 (ene. de 2018).

¹⁸ Leo Breiman. “Random Forests”. En: *Machine Learning* 45.1 (1 de oct. de 2001), págs. 5-32. DOI: 10.1023/A:1010933404324.

¹⁹ George Grekousis. “Artificial neural networks and deep learning in urban geography: A systematic review and meta-analysis”. En: *Computers, Environment and Urban Systems* 74 (2019), págs. 244-256. DOI: <https://doi.org/10.1016/j.compenvurbsys.2018.10.008>.

Series temporales Según Peter J. Brockwell y Richard A. Davis ²⁰, las series temporales se definen como un conjunto de observaciones X_t , cada una registrada en un punto del tiempo específico t , es decir, una serie de tiempo es una secuencia de datos observados en intervalos ya sea discretos o continuos. Es importante tener en cuenta que las series de tiempo pueden tener diferentes patrones y características, como tendencias, estacionalidad, ciclos y ruido. El análisis de estas características es fundamental para comprender y modelar adecuadamente los datos de la serie de tiempo.

2.5.1. Modelos de Predicción en series temporales En el ámbito de la predicción en series temporales, existen diversas técnicas y modelos diseñados para capturar patrones temporales y realizar pronósticos precisos. A continuación, se describen algunos de los modelos más comunes utilizados en este contexto:

Autoregressive Integrated Moving Average (ARIMA) es una técnica clásica de predicción en series temporales que combina componentes autoregresivos, de media móvil e integrados. Es particularmente efectivo para datos estacionarios y permite modelar la tendencia y la estacionalidad. También permite analizar datos sin el requerimiento de independencia de tiempo de errores²¹.

Seasonal-Trend decomposition using LOESS (STL) es un enfoque que descompone una serie temporal en tres componentes principales: estacionalidad, tendencia y residuos locales. Esta descomposición facilita la modelación individual de cada componente, permitiendo una predicción más precisa²².

Long Short-Term Memory (LSTM) Las redes neuronales recurrentes, como LSTM, son modelos de aprendizaje profundo especialmente diseñados para capturar dependencias a largo plazo en series temporales. Son adecuadas para datos no lineales y secuenciales, logrando un rendimiento destacado

²⁰ Peter J. Brockwell y Richard A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. New York, NY: Springer, 1991. DOI: 10.1007/978-1-4419-0320-4.

²¹ Brian K. Nelson. "Time Series Analysis Using Autoregressive Integrated Moving Average (ARIMA) Models." En: *Academic Emergency Medicine* 5 (28 de jun. de 2008).

²² Jean E. McRae Robert B. Cleveland William S. Cleveland e Irma Terpenning. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess". En: *Journal of Official Statistics* 6 (mar. de 1990).

en contextos temporales complejos²³.

Prophet es un modelo desarrollado por Facebook diseñado para la predicción de series temporales con fuertes componentes estacionales. Utiliza una descomposición aditiva basada en tendencias y estacionalidad para realizar pronósticos precisos.²⁴

2.6. APLICACIÓN DE MODELOS ESPACIALES Y TEMPORALES

En la incansable búsqueda de metodologías avanzadas que aborden tanto la predicción espacial como temporal, diversos estudios han explorado terrenos más allá de los límites de campos específicos. Este compendio de conocimientos no solo proporciona una visión integral de las capacidades de predicción, sino que también sirve como plataforma fundamental para la investigación actual, centrada en anticipar emergencias médicas.

Al adentrarnos en uno de los pilares esenciales de esta investigación, el análisis de fenómenos geoespaciales, el estudio ²⁵ emerge como un faro en la comprensión de casos complejos. Resaltando la insuficiencia de los modelos lineales convencionales, este estudio aborda de manera exhaustiva la necesidad de un análisis espacial detallado para evaluar la presencia de dependencia espacial entre deslizamientos. La metodología propuesta combina técnicas de aprendizaje automático, como la red neuronal parcialmente recurrente, con herramientas de estadística espacial, como la regresión geográficamente ponderada. Este enfoque audaz marca un hito significativo hacia la mejora de la precisión predictiva en el espacio.

²³ Lee G. Le X-H. Ho HV. y Jung S. “Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting.” En: *Water* 11 (5 de jul. de 2019).

²⁴ Valagolam D. Shen J. y McCalla S. “Prophet forecasting model: a machine learning approach to predict the concentration of air pollutants (PM2.5, PM10, O3, NO2, SO2, CO) in Seoul, South Korea”. En: *PeerJ* (15 de sep. de 2020).

²⁵ Andrés Camilo Florez García y José Nelson Pérez Castillo. “Técnicas para la predicción espacial de zonas susceptibles a deslizamientos”. En: *Avances Investigación en Ingeniería* 16.1 (21 de ago. de 2019). Number: 1, págs. 20-48. DOI: 10.18041/1794-4953/avances.1.5188.

Simultáneamente, la investigación ²⁶ nos sumerge en la complejidad temporal y espacial de las variaciones solares. Al destacar la importancia de la geoestadística y las redes neuronales artificiales en este contexto, este estudio sugiere que la consideración simultánea de los aspectos espaciales y temporales es esencial para la predicción de emergencias médicas.

La metodología detallada en este segundo estudio se basa en la evaluación de la disponibilidad de datos meteorológicos en tiempo cuasi real en el territorio español. Se caracterizan las redes de estaciones meteorológicas identificadas, se estudian y evalúan los métodos de estimación de la irradiancia solar global a intervalos de 15 minutos, y se experimenta con modelos basados en Redes Neuronales Artificiales (RNAs). La inclusión de la componente espacial en las entradas de las RN y la relación entre el horizonte espacial y temporal en la predicción de la irradiancia solar a corto plazo se exploran detalladamente.

En la exploración de estos métodos para la predicción espacial y temporal el estudio Red neuronal como herramienta de predicción de variables climáticas en Castilla y León "se centra en la aplicación de Redes Neuronales Artificiales para predecir en un fenómeno de alta complejidad como lo son las variables climáticas en una región específica. La metodología detallada en este estudio se entrelaza con la rigurosa evaluación de la disponibilidad de datos meteorológicos en tiempo cuasi real en el territorio español. Así como en el estudio anterior se caracterizaron las redes de estaciones meteorológicas, aquí se lleva a cabo una evaluación similar, destacando la importancia de comprender y aprovechar la infraestructura ya existente para el análisis climático.

Además, al estudiar y evaluar los métodos de estimación de la irradiancia solar global a intervalos de 15 minutos, el estudio sobre las variables climáticas refuerza la necesidad de una consideración temporal detallada. La relación entre el horizonte espacial y temporal en la predicción de la irradiancia solar a corto plazo, detalladamente explorada en ambos estudios, subraya la sinergia crítica entre los factores espaciales y temporales en la anticipación de eventos climáticos.

La experimentación con modelos basados en RNAs, destacada en ambos estudios, revela no solo la

²⁶ Federico Vladimir Gutiérrez Corea. "Predicción espacio-temporal de la irradiancia solar global a corto plazo en España mediante geoestadística y redes neuronales artificiales". <http://purl.org/dc/dcmitype/Text>. Universidad Politécnica de Madrid, 2014.

versatilidad de esta herramienta en diferentes contextos, sino también su capacidad para modelar complejidades espaciales y temporales. La inclusión de la componente espacial en las entradas de las RNAs, vista como un componente clave en el estudio anterior sobre deslizamientos, encuentra su correlato en este estudio climático, consolidando así la importancia de considerar la interacción dinámica entre aspectos espaciales y temporales en nuestra búsqueda por la predicción efectiva.

En el incesante empeño por explorar metodologías avanzadas que aborden la predicción espacial y temporal, se dirige la atención hacia el estudio de José Luis Labajo Salazar²⁷. Este análisis se dedica exclusivamente a la predicción de delitos, proporcionando una perspectiva única al abordar de manera detallada los complejos factores que influyen en la ocurrencia de eventos delictivos en entornos urbanos. La metodología de este estudio se distingue por su enfoque espaciotemporal, centrándose en la interacción dinámica entre los elementos geoespaciales y las variaciones temporales. A diferencia de otros enfoques, se lleva a cabo una exhaustiva recopilación de datos urbanos y demográficos para entender la estructura socioeconómica y geográfica de Buenos Aires. Se aplican técnicas avanzadas de análisis espacial y temporal, sin atarse a un algoritmo específico, para modelar la evolución de los patrones delictivos en diferentes áreas de la ciudad a lo largo del tiempo.

Cambiando de escenario, se sumerge en el análisis J.Díaz-Hierro²⁸. Este estudio, en un ámbito distinto, pero igualmente crítico, se centra en las necesidades de anticipación en el sector de la salud. La evaluación meticulosa de modelos de series temporales es clave en este análisis, destacando la importancia de prever con precisión la demanda de servicios sanitarios y adaptarse a cambios estacionales y eventos extraordinarios.

Al vincular estos dos estudios con el recorrido previo, se encuentra un hilo conductor: la esencial interacción entre los factores espaciales y temporales en la predicción. Así como el estudio sobre deslizamientos y variables climáticas enfatizaba la necesidad de considerar ambas dimensiones, estos estudios ofrecen una visión enriquecedora de cómo esta interacción se manifiesta en contextos urbanos y de salud.

²⁷ Rafael Zambrano. “Un enfoque espaciotemporal para la predicción de delitos en la ciudad de buenos aires”. En: *Revista de investigación en modelos matemáticos aplicados a la gestión y la economía* 2.7 (13 de nov. de 2021).

²⁸ J. Díaz-Hierro et al. “Evaluación de modelos de series temporales para la previsión de la demanda de emergencias sanitarias”. En: *Emergencias (St. Vicenç dels Horts)* (2012), págs. 181-188.

En este trayecto, se ha explorado desde la predicción de deslizamientos hasta la anticipación de la demanda sanitaria y delitos urbanos. Cada paso acerca más a una comprensión holística de cómo las metodologías avanzadas pueden ofrecer soluciones precisas y contextualmente relevantes.

3. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA

En la actualidad, los servicios de salud son esenciales para garantizar el bienestar de la población y un desarrollo constante de las comunidades. En consecuencia, los servicios de salud deben asegurar, en la medida de los recursos, una atención con respuesta oportuna y gran cobertura. Las responsabilidades que implica prestar el servicio deben cumplirse basadas en las funciones esenciales de la salud y garantizarse en toda la ruta de atención ver figura 3. Así, la calidad, oportunidad y efectividad de la atención prestada debe respaldar una ejecución integral del plan obligatorio de salud. En general, la eficiencia en el servicio de salud es fundamental para preservar la vida y la salud de los usuarios, bien sea su ingreso debido a una emergencia o a una remisión del servicio.

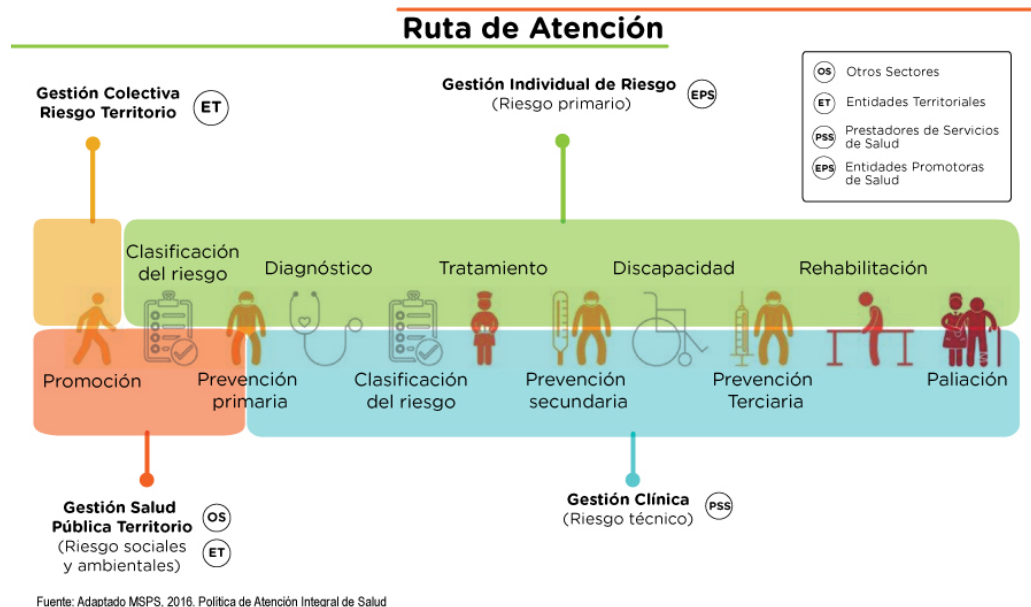


Figura 3. Ruta de atención. Tomado de la página web del ministerio de salud y protección social de Colombia. extraído de la página del Ministerio de Salud y Protección Social²⁹

La atención ante emergencias médicas es parte fundamental de los servicios de salud que se prestan a los usuarios y un gran factor en su bienestar. En el marco de las emergencias médicas enfocadas en la atención prehospitalaria existe una necesidad de conocimiento frente a los focos geográficos donde

se generan la mayor cantidad de emergencias, debido a que la administración y coordinación de estas entidades afronta diariamente diversas problemáticas tales como: limitación en los recursos físicos, de personal y factores administrativos cruciales. Tomando en cuenta las anteriores necesidades y limitaciones expuestas se debe abordar la problemática desde la representación misma de estos eventos para que los actores administrativos puedan basar sus estrategias y análisis en un marco general, el cual se debe poder reproducir, escalar y predecir.

La actual disposición de los servicios prehospitalarios se encuentra contenida en el Sistema de Emergencias Médicas (SEM), cuya reglamentación está descrita en la Resolución 926 de 2017 del Ministerio de Salud¹. Este sistema tiene el propósito de responder de manera oportuna a las víctimas de enfermedad, accidentes de tránsito, traumatismos o paros cardiorrespiratorios que requieran atención médica de urgencias en lugares públicos o privados. El SEM coordina los diversos actores del Sistema General de Seguridad Social en Salud para garantizar una respuesta oportuna a las víctimas de accidentes y enfermedades generales que requieran atención inmediata. Además, utiliza mecanismos de notificación, respuesta y gestión de solicitudes para la optimización de recursos. No obstante, no existe un mecanismo de apoyo para identificar patrones espaciales, temporales y clínicos asociados a las emergencias, por lo que se espera que el análisis de las emergencias médicas, y modelos de predicción basados en aprendizaje máquina permitan acercarse al cumplimiento del objetivo principal de estas organizaciones SEM, el cual es “llevar el paciente indicado, al hospital adecuado en el momento oportuno”³⁰.

Un ejemplo de aplicación para los modelos de predicción de emergencias podría ser la optimización de las bases de atención inmediata (puntos geográficos estratégicos donde se sitúan las ambulancias). La ubicación de las bases requiere un análisis de las emergencias dentro de la zona de influencia. Esta actividad puede ser apoyada por modelos de predicción de las emergencias y el uso de algoritmos de optimización. Estos algoritmos deberán tener en cuenta factores geográficos, franjas de horas pico, afluencia de tráfico, tiempos de desplazamiento, entre otros. Además, se podría apoyar la toma de de-

³⁰ Ministerio de salud y protección social. *Sistema de emergencias médicas SEM*. Sistema de emergencias medicas-SEM. 29 de dic. de 2019. URL: <https://www.minsalud.gov.co/salud/PServicios/Paginas/Sistema-de-emergencias-medicas-SEM.aspx> (visitado 23-08-2023).

cisiones de planeación urbanas, como por ejemplo en la planificación del tránsito vehicular, dando lugar a planes de movilidad que tengan en cuenta factores asociados a focos de accidentabilidad.

Teniendo en cuenta la situación expuesta, este proyecto se propone implementar un modelo de predicción que permita analizar y estimar focos de emergencias médicas en los que se requiera atención prehospitalaria mediante un enfoque espacial y temporal. Este modelo propiciará información clave de las características de cada emergencia tales como descripción geográfica aparente, marcado temporal y características descriptivas del evento.

Esta experiencia no solo brindará oportunidades para aplicar técnicas avanzadas de gestión y tratamiento de datos, utilizando diversas fuentes de información como base, sino que también representará un desafío significativo. Este desafío se traduce en la generación de predicciones confiables sobre la ocurrencia de emergencias en un horizonte temporal amplio y en la identificación de características críticas que influyen en la toma de decisiones.

Las representaciones espaciales y el uso de variables temporales presentan retos adicionales, como la complejidad en la representación de datos espaciales, la definición de la resolución del modelo y las dependencias temporales y espaciales de las variables, entre otros aspectos. Superar la dificultad inherente a la identificación de metodologías eficaces para la predicción de emergencias y conservar la naturaleza espacial y temporal de los datos requiere la creación de una estrategia no lineal basada en aprendizaje automático.

Esta estrategia no solo permitirá obtener predicciones fiables sobre la afluencia de emergencias, sino que también contribuirá significativamente a la evaluación de representaciones temporales y espaciales de estos eventos. La capacidad de adaptar el modelo a un espacio definido y específico garantiza la generación de predicciones adaptadas a la naturaleza espacial única de la zona de estudio. En este contexto, el proyecto se posiciona como una herramienta beneficiosa para la toma de decisiones en sistemas de respuesta inmediata, ofreciendo información esencial para el despliegue eficiente de servicios de ambulancia y la gestión efectiva de situaciones críticas.

4. OBJETIVOS

Objetivo general

Desarrollar un modelo de aprendizaje automático basado en datos que presentan variables espaciales y temporales para predecir la ocurrencia de emergencias médicas en una zona geográfica determinada.

Objetivos específicos

- Construir una base de datos asociada a individuos y/o organizaciones que conformen el sistema de emergencias médicas enfocado a la atención prehospitalaria en una región geográfica específica.
- Formular una representación espacial y temporal para el análisis de las emergencias médicas.
- Implementar un modelo predictivo que utilice la representación espacial y temporal para estimar en un instante de tiempo futuro la probabilidad de ocurrencia de una emergencia médica en una ubicación específica.
- Evaluar la efectividad del modelo propuesto en la predicción de emergencias médicas, bajo diversos valores en la configuración de los parámetros, para una zona geográfica determinada.

5. MATERIALES

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

5.1. ZONA DE ESTUDIO

La exposición de esta zona como parte central del estudio permite dar un contexto de los datos utilizados y como se aplicará el modelo predictivo, permitiendo considerar las características geospaciales y demográficas que pueden influir en el comportamiento del fenómeno.

5.1.1. Ubicación geográfica Bucaramanga es un municipio del departamento de Santander ubicado en Colombia, más específicamente en 7.118611° , -73.116111° . Esta ubicación fue seleccionada debido a su importancia en términos de densidad poblacional, infraestructura médica y variabilidad en la geografía local. La figura 4 puede dar una relación visual a la zona de referencia.

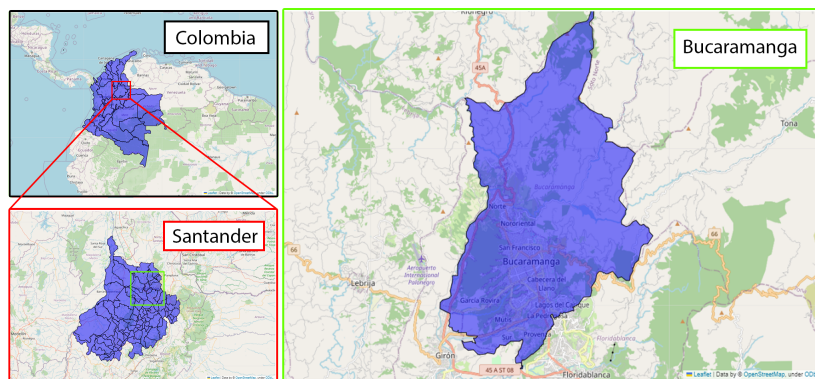


Figura 4. Área de estudio. Creación propia.

5.1.2. Contexto Geoespacial El área de estudio expuesta se caracteriza por poseer factores geográficos particulares que influyen directamente en la distribución y comportamiento de las emergencias. La climatología y topografía desempeña un rol particular en la dinámica de dichos eventos y expone retos en la respuesta de estos. La región municipal de Bucaramanga abarca 165 kilómetros cuadrados y se encuentra a una altitud promedio de 959 metros sobre el nivel del mar. Los pisos térmicos varían con 55 kilómetros cuadrados de clima cálido, 100 kilómetros cuadrados de clima medio y 10 kilómetros cuadrados de clima frío. La temperatura media es de 23°C, y la precipitación anual promedio alcanza los 1.041 mm. La topografía muestra una mezcla de características, con aproximadamente un 15 % de área plana, un 30 % ondulado y el 55 % restante en terreno quebrado. Tres prominentes cerros, Morro Rico, Alto de San José y El Cacique, definen la silueta de la región. En cuanto a la hidrografía, los ríos principales son el Río de Oro y el Suratá, acompañados por una red de quebradas que incluye La Flora, Tona, La Iglesia, Quebrada Seca, Cacique, El Horno, San Isidro, Las Navas y La Rosita. Esta información se extrajo de AMB³¹.

5.1.3. Datos Demográficos 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. Los índices de juventud, envejecimiento y dependencia demográfica correspondientes son 25.78, 54.56 y 42.49 ³².

5.1.4. Instituciones de interés

DTB La Dirección de Tránsito de Bucaramanga es una entidad pública descentralizada del orden municipal, la cual fue creada con el propósito de organizar y controlar todo lo relacionado con el tránsito

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

³² Departamento Administrativo Nacional de Estadística (DANE). *Censo Nacional de Población y Vivienda 2018*, DANE. 10 de mayo de 2019.

dentro de Bucaramanga y velar por el cumplimiento de las disposiciones legales sobre la materia³³.

CRUE El Centro Regulador de Urgencias y Emergencias es una unidad de carácter operativo no asistencial, responsable de coordinar y regular en el territorio de su jurisdicción, el acceso a los servicios de urgencias y la atención en salud de la población afectada en situaciones de emergencia o desastres³⁴.

Instituciones prestadoras de servicios de salud Dado que la atención de las emergencias médicas posee un sector privado involucrado directamente, la predicción de focos de este fenómeno podrá ayudar en la ubicación estratégica de las diversas bases de atención dentro del área metropolitana, favoreciendo el tiempo de respuesta y la optimización de los recursos de estas entidades. La posibilidad de prever la disposición de las emergencias podría permitir evitar situaciones tales como “La guerra del centavo” producto de la lucha por cobros de los Seguros Obligatorios contra Accidentes de Tránsito (SOAT)³⁵.

Según el Registro Especial de Prestadores de Servicios de Salud (REPS) en Colombia existen alrededor de 290 entidades habilitadas para prestar servicios prehospitales. En Santander, el departamento en el cual se ubica la zona de estudio, existen 20 de estas entidades las cuales cuentan con la habilitación para prestar estos servicios³⁶, por lo tanto tener en cuenta estas entidades en el planteamiento de esta investigación es crucial.

³³ Dirección de Tránsito de Bucaramanga. *Nuestra Entidad*. DTB - Dirección de Tránsito de Bucaramanga. URL: <https://transitobucaramanga.gov.co/dtb/quienes-somos/nuestra-entidad/> (visitado 31-08-2023).

³⁴ Ministerio de Salud y Protección Social. *Resolución Número 1220 de 2010*. 8 de abr. de 2010.

³⁵ Semana. “Ambulancias sin límite: la ‘guerra del centavo’ continúa y está dejando víctimas mortales”. En: *Semana.com Últimas Noticias de Colombia y el Mundo* (18 de jun. de 2022). Section: Enfoque.

³⁶ Ministerio de Salud y Protección Social. *Registro Especial de Prestadores de Servicios de Salud (REPS)*. 10 de dic. de 2023.

5.2. DESCRIPCIÓN DE FUENTES DE DATOS

En esta sección se describirán los procedimientos, datos de interés, métodos de tratamiento, entre otras consideraciones previstas en el tratamiento y manipulación de los datos.

5.2.1. Registro de emergencias médicas y accidentes de tránsito Los datos proporcionados por las instituciones de interés se solicitan por medio de los canales de PQRs de las mismas. A continuación se describirán cada uno de los formatos y características de cada uno de los conjuntos de datos.

Los datos proporcionados por el Centro regulador de Urgencias y Emergencias son registros de emergencias médicas ocurridas en el área metropolitana de Bucaramanga con un intervalo de tiempo de dos años (2021-2022), estos datos poseen las siguientes características: Fecha de atención, Nombre del Usuario / Persona quien reporta, Lugar del evento, Comuna, Edad, Sexo, Hora de Reporte, Clasificación del TRIAGE, IPS Ambulancia quien atiende el caso, Base ambulancia asignada, Hora de llegada de la ambulancia, Hora de llegada a la IPS, IPS de recepción del usuario y Observaciones.

A su vez la Dirección de Tránsito de Bucaramanga proporciona registros de accidentes de tránsito con un periodo de tiempo de alrededor de 6 años (2017 a 2022). Estos registros poseen las siguientes características: número de plantilla, día de la semana, día del mes, mes, año, fecha, gravedad, moto, bicicleta, otro, dirección, barrio, hora, diurno/nocturno y comuna.

Dado el enfoque de este trabajo de investigación se optó por tomar de estos dos conjuntos de datos las variables representativas del espacio y las variables temporales del mismo, estas son: la fecha de ocurrencia de cada evento y su respectiva dirección.

5.2.2. Datos Geospaciales Las fuentes de datos geospaciales son:

Marco Geoestadístico Nacional (MGN) Es un marco de referencia espacial obligatorio establecido por el Departamento Administrativo Nacional de Estadística (DANE) en Colombia. Este marco se utiliza para el desarrollo del proceso estadístico y facilita las consultas geográficas de los resultados del Censo Nacional de Población y Vivienda³⁷.

El MGN proporciona una estructura jerárquica de niveles geográficos que van desde departamentos y municipios hasta manzanas, con el fin de organizar y representar la información geográfica del país. Cada nivel geográfico contiene atributos y variables censales asociadas, lo que permite analizar la distribución de la población y otras características a diferentes escalas espaciales.

En esta investigación se tiene como interés dos secciones del MGN las cuales son:

MGN nivel municipio se refiere a la distribución de entidades territoriales fundamentales en la división político-administrativa del Estado, con autonomía política, fiscal y administrativa dentro de los límites que le señalan la Constitución y las leyes de la República. En este nivel, se encuentran 1.101 municipios. Bogotá, D. C. se codifica como municipio, junto con las 20 áreas no municipalizadas y San Andrés Isla, lo que da un total de 1.122 polígonos en el nivel geográfico³⁷.

En el caso de esta investigación los datos utilizados ³⁸ están compuestos por 12 atributos geográficos propios del nivel municipal. Estos atributos incluyen el código DANE del departamento, el código DANE del municipio, código DANE concatenado del departamento y municipio, el nombre de la entidad territorial departamental, el nombre de la entidad territorial municipal, el año del documento creación del municipio, el área del departamento en kilómetros cuadrados, el símbolo polígono del municipio, el año de vigencia DANE y tipo de municipio según la división político-administrativa de Colombia (DIVIPOLA).

Estos datos proporcionarán una estructura detallada (ver figura 5) permitiendo así una validación

³⁷ Departamento Administrativo Nacional de Estadística (DANE). *Uso del marco geoestadístico nacional*. Sep. de 2020.

³⁸ Departamento Administrativo Nacional de Estadística (DANE). *Marco Geoestadístico Nacional Nivel Municipio 2022*. Ver. MGN2022-Nivel Municipio. 7 de dic. de 2022.

espacial de las emergencias médicas y otros análisis dentro del marco de la investigación.

MGN sección urbana Hace referencia a la división cartográfica creada por el DANE con fines estadísticos, conformada por secciones urbanas. Es la mayor división o máximo nivel de agregación definido dentro del perímetro censal de las cabeceras municipales y centros poblados³⁷.

En este nivel se seleccionaron los datos ³⁹ para hacer parte de la investigación, ya que tiene identificados 11,570 polígonos que representan sectores urbanos. Cada uno de estos polígonos está caracterizado por 8 atributos geográficos específicos de este nivel. Estos atributos abarcan información como el código DANE del departamento, el código DANE del municipio, el código DANE concatenado del departamento, municipio y centro poblado, el código DANE concatenado del departamento, municipio, clase, sector rural, sección rural, centro poblado y sector urbano, así como el código DANE concatenado del departamento, municipio, clase, sector rural, sección rural, centro poblado, sector urbano y sección urbana. Además, se incluyen datos como el área del sector urbano en kilómetros cuadrados, el perímetro del sector urbano en kilómetros y el símbolo del polígono que representa la sección urbana.

La gráfica correspondiente a la figura 5 ofrece una perspectiva detallada de la zona de estudio, destacando dos elementos cruciales. En la parte izquierda se enfoca en la visualización del polígono correspondiente al municipio de interés, proporcionando un análisis exhaustivo de su división política interna. Esta representación permite explorar la distribución de áreas específicas dentro del municipio, destacando características políticas y geográficas a nivel local. En contraste, la sección derecha de la gráfica exhibe un seccionamiento urbano irregular, delineando de manera precisa los límites y características de los sectores urbanos dentro del área del municipio. Esta dualidad en la visualización brinda una comprensión integral de la estructura territorial, con un enfoque particular en el municipio de estudio y los detalles específicos de su entorno urbano.

³⁹ Departamento Administrativo Nacional de Estadística (DANE). *Marco Geoestadístico Nacional Sección Urbana 2022*. Ver. MGN2022-Nivel Sección urbano. 7 de dic. de 2022.

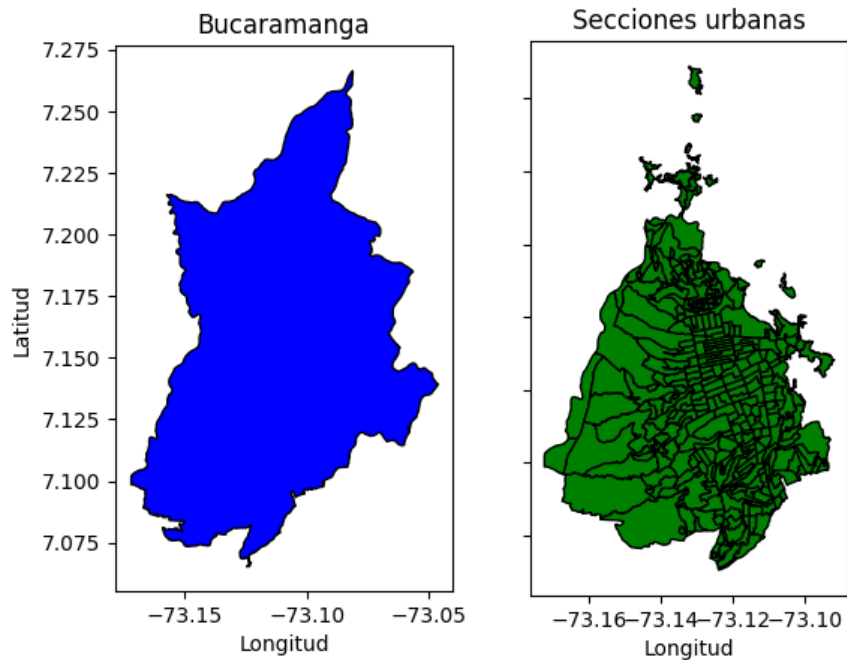


Figura 5. MGN de división política del municipio de Bucaramanga y sus secciones urbanas. Creación propia con geopandas⁴⁰ y matplotlib.⁴¹

OSMnx es una herramienta que permite a los académicos y profesionales urbanos adquirir, construir, analizar y visualizar datos de redes de calles complejas. Con OSMnx, se pueden descargar fácilmente límites políticos y huellas de edificios, construir datos de redes de calles a partir de OpenStreetMap⁴², corregir la topología de la red, guardar redes de calles en varios formatos de archivo y analizar redes de calles utilizando una variedad de métricas y medidas⁴³. En la presente investigación se utiliza OSMnx de manera recurrente en el planteamiento de la representación espacial y temporal de las emergencias médicas y accidentes de tránsito.

La figura 18 proporciona un ejemplo visual de este proceso, destacando las intersecciones clave y la red

⁴² OpenStreetMap contributors. *Planet dump* retrieved from <https://planet.osm.org>. <https://www.openstreetmap.org>. 2017.

⁴³ Geoff Boeing. "OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks". En: *Computers, Environment and Urban Systems* 65 (1 de sep. de 2017), págs. 126-139. DOI: 10.1016/j.compenvurbsys.2017.05.004.

vial que caracteriza la topografía urbana de Bucaramanga. Gracias a la funcionalidad de OSMnx, se logra una representación detallada y actualizada de las vías, incluyendo su disposición y conectividad. Este enfoque no solo contribuye a la comprensión espacial de la red vial, sino que también sienta las bases para análisis más profundos, como la evaluación de la eficiencia del flujo de tráfico, la identificación de puntos críticos y la planificación de medidas para mejorar la movilidad urbana. La aplicación estratégica de OSMnx, evidenciada en este gráfico, demuestra su capacidad para respaldar investigaciones que requieren una representación espacial precisa y actualizada de la infraestructura vial en entornos urbanos específicos.

5.2.3. Datos demográficos Los principales datos poblacionales utilizados fueron: el número de personas, el conteo de personas entre 0 a 9 años, el conteo de personas entre 10 a 19 años, el conteo de personas entre 20 a 29 años, el conteo de personas entre 30 a 39 años, el conteo de personas entre 40 a 49 años, el conteo de personas entre 50 a 59 años y el conteo de personas entre 80 y mas años. Estos datos fueron extraídos del censo nacional de población y vivienda de Colombia realizado en el 2018³², a través del MGN integrado⁴⁴.

5.2.4. Datos climáticos Los datos hidrometeorológicos utilizados en el estudio son: la temperatura (°C), precipitación (mm), dirección del viento (°), velocidad del viento (m/s) y presión atmosférica (hPa). Las fuentes de datos que se utilizaron fueron:

Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM Para enriquecer la investigación, se recuperaron datos cruciales de precipitación en milímetros provenientes de dos estaciones meteorológicas ubicadas estratégicamente en el municipio de Bucaramanga⁴⁵. La disposición de estos datos se presenta visualmente en la figura 6 que ilustra la precipitación de las estaciones hidrológicas en el área metropolitana de Bucaramanga.

⁴⁴ Departamento Administrativo Nacional de Estadística (DANE). *Marco Geoestadístico Nacional Sección Urbana integrado al CNPV2018*. Ver. Versión MGN_ANM2018-Sección Urbana. 27 de dic. de 2021.

⁴⁵ Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM. *Datos de precipitación de pluviómetro convencional en Santander, Bucaramanga*. 16 de dic. de 2023.

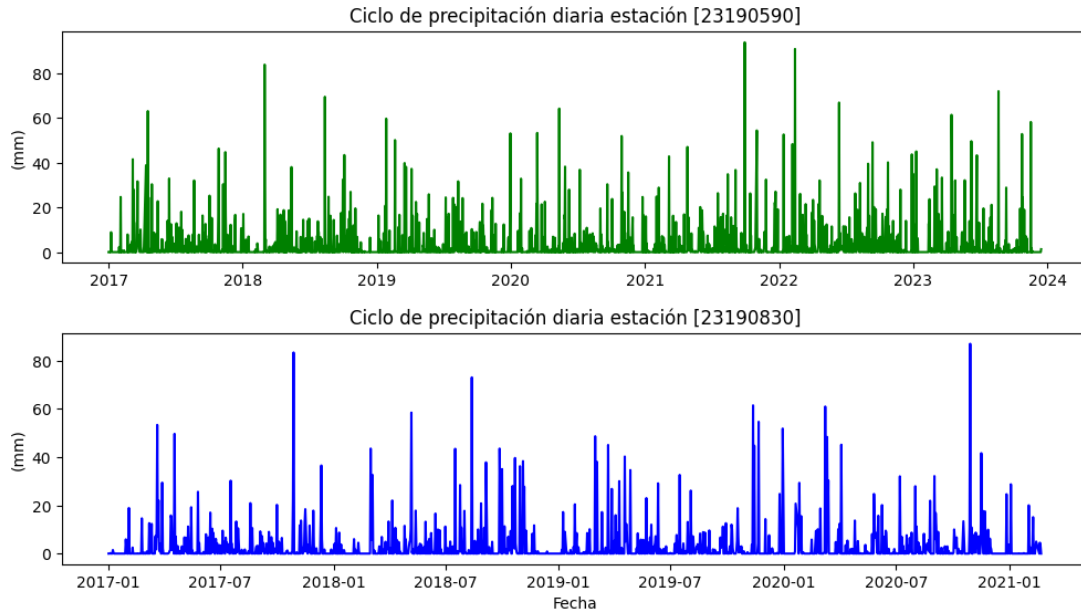


Figura 6. Precipitación diaria en estaciones meteorológicas de Bucaramanga. Creación propia con pandas y matplotlib.

En la parte superior de la gráfica, se detallan los registros de la estación denominada Floresta con el código de estación 23190590. Esta estación está situada en la latitud 7.090277778, longitud -73.12388889, y una altitud de 833 metros sobre el nivel del mar. Los datos recopilados desde el 31 de diciembre de 2016 hasta la fecha actual proporcionan una valiosa serie temporal para analizar patrones y variaciones en la precipitación en esta ubicación específica.

En la parte inferior de la gráfica, se exhiben los datos de la estación Bucaramanga IDEAM. Esta estación, ubicada en la latitud 7.128333333, longitud -73.11833333, y a una altitud de 1001 metros sobre el nivel del mar, ha registrado información desde el 31 de diciembre de 2016 hasta el 21 de febrero de 2021, fecha en la que suspendió su funcionamiento. Estas series temporales ofrecen una visión detallada de las condiciones hidrometeorológicas en Bucaramanga y su área metropolitana, permitiendo análisis y comparaciones significativas para comprender la variabilidad de la precipitación en el tiempo y el espacio.

meteostat La librería Meteostat Python emerge como una herramienta eficiente para acceder de manera sencilla a datos meteorológicos y climáticos, aprovechando la potencia de la libreríaPandas⁴⁶. Los registros y estadísticas históricas obtenidos mediante esta librería se originan en la interfaz de datos masivos de Meteostat⁴⁷, consolidando información recopilada de diversas fuentes públicas, en su mayoría gubernamentales. Entre estas fuentes destacan servicios meteorológicos nacionales como la Administración Nacional Oceánica y Atmosférica (NOAA) y el servicio meteorológico nacional de Alemania (DWD).

La robustez de Meteostat Python se evidencia en su capacidad para aglutinar datos provenientes de estas fuentes heterogéneas, proporcionando a los usuarios un acceso integral a información meteorológica de alta calidad. Las variables recuperadas mediante esta librería no solo incluyen registros básicos, sino que abarcan un espectro completo de parámetros meteorológicos cruciales para análisis detallados. La utilización de Meteostat Python se revela como un recurso valioso en la investigación, permitiendo la exploración y comprensión profunda de las condiciones climáticas históricas con un enfoque basado en datos confiables y diversificados.

Para complementar el análisis se han creado dos gráficas que presentan series temporales de datos meteorológicos significativos desde enero de 2021 hasta la actualidad. En la primera gráfica, se visualiza la evolución de la temperatura media, velocidad del viento y presión atmosférica a lo largo del tiempo. Cada serie temporal proporciona una comprensión detallada de las variaciones diarias de estas variables meteorológicas clave, permitiendo identificar patrones estacionales, tendencias y eventos anómalos véase la figura 7.

⁴⁶ The pandas development team. *pandas-dev/pandas: Pandas*. Ver. latest. Feb. de 2020. DOI: 10.5281/zenodo.3509134.

⁴⁷ Christian Lamprecht. *Meteostat*. Ver. 1.6.7. Germany, 28 de sep. de 2023.

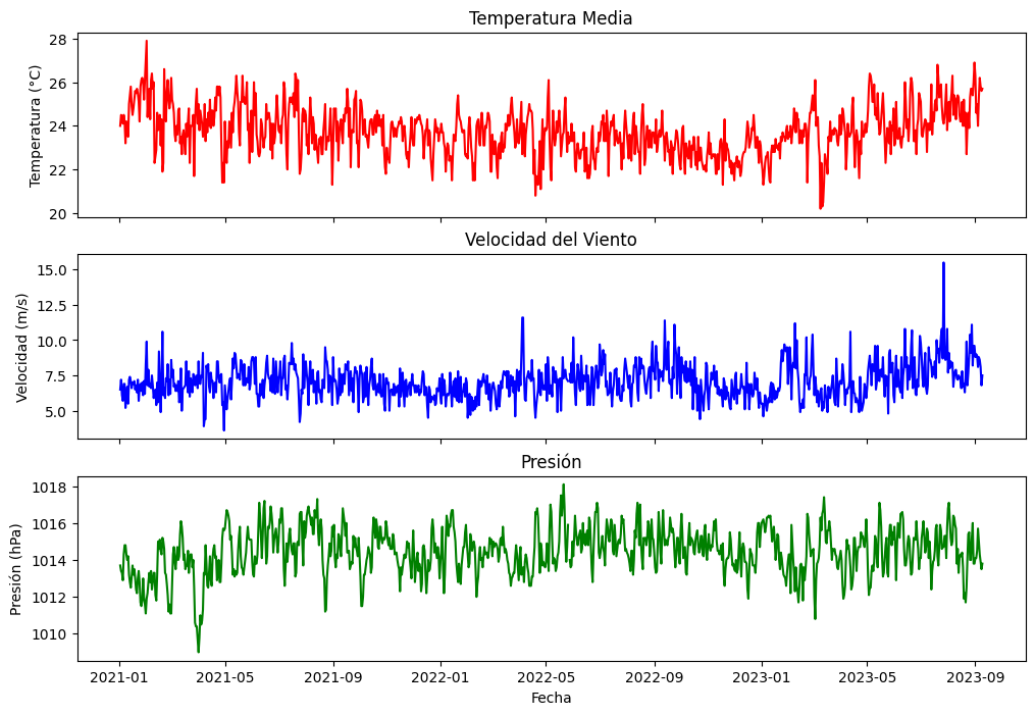


Figura 7. Histórico de variables climáticas en el área metropolitana de Bucaramanga. Creación propia.

En la segunda gráfica, de tipo radial, se presenta la dirección del viento en grados y la frecuencia asociada a cada dirección. Esta representación visual facilita la comprensión intuitiva de la distribución del viento en la zona de estudio, identificando las direcciones predominantes y su frecuencia relativa. La disposición circular de esta gráfica proporciona una representación efectiva de la variabilidad direccional del viento. Véase la figura 8.

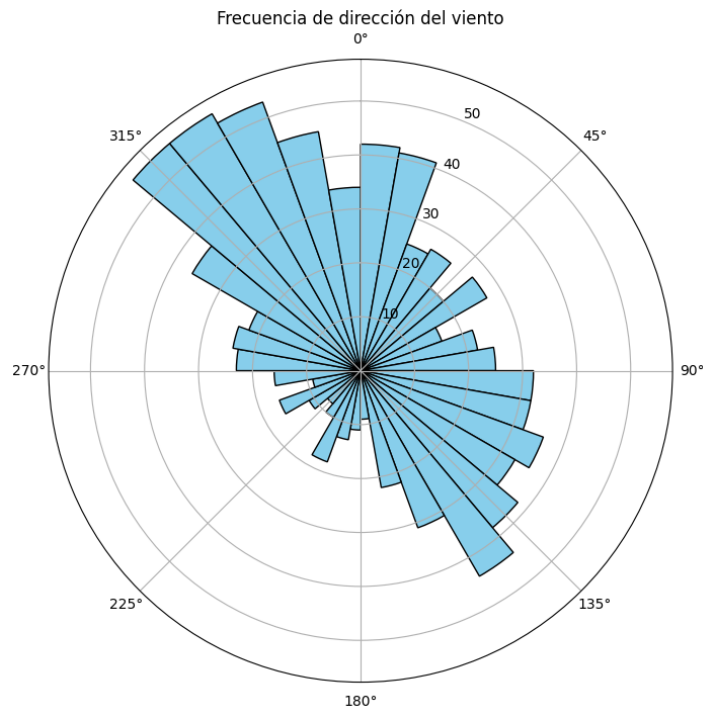


Figura 8. Frecuencia de la dirección del viento en grados. Creación propia.

Ambas visualizaciones buscan proporcionar una perspectiva completa y accesible de las condiciones meteorológicas en la zona de estudio, facilitando la interpretación de patrones climáticos y la identificación de comportamientos específicos asociados a las variables consideradas.

5.2.5. Datos temporales En el marco de este estudio se han recopilado y estructurado datos temporales esenciales que abarcan el extenso periodo desde enero de 2017 hasta diciembre de 2023. Estos datos formarán parte integral del modelo, proporcionando información detallada sobre días festivos, feriados y días normales en Colombia durante este intervalo temporal. La representación gráfica de estos datos, desarrollada con las librerías Pandas, Calmap⁴⁸ y Matplotlib, contiene dos categorías significativas de festividades: las reconocidas oficialmente por el estado, que corresponden a días de descanso remunerado; y otro grupo de días, denominados conmemoraciones sociales, que, aunque no son

⁴⁸ Martijn Vermaat y otros. *Calmap: Calendar heatmaps from Pandas time series data*. Ver. 0.0.11. 31 de jul. de 2023.

considerados días de descanso, son celebrados y pueden estar relacionados con fenómenos de interés en el estudio. Entre estas conmemoraciones sociales se encuentran celebraciones como el Día Internacional de la Mujer, el Día del Padre, el Día de la Madre, el Día de San Valentín, Halloween, el Día del Año Nuevo, el día del Amor y la Amistad, entre otras. Esta clasificación detallada de festividades ofrece una comprensión más profunda de la variabilidad temporal en eventos relacionados y permitiendo al modelo poder relacionarlas con el comportamiento de las emergencias médicas. En la gráfica correspondiente a la figura 9 todas las festividades se presentan de manera uniforme, destacando el día indexado en alguna de las dos categorías mencionadas con color negro. Esta distinción visual resalta la relevancia de estos eventos específicos en el contexto temporal del estudio.

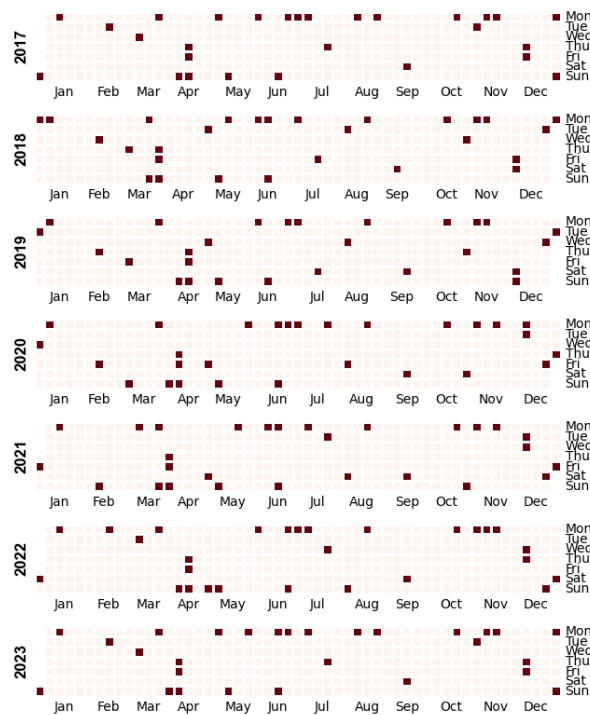


Figura 9. Días de importancia social. Creación propia con pandas y calmap.

5.2.6. Datos de tráfico y movilidad

Google Maps Platform Considerando la naturaleza diversa y multifactorial de las emergencias médicas, se ha explorado la posibilidad de establecer conexiones entre estos eventos y los accidentes

de tránsito, así como otros sucesos cotidianos en la movilidad urbana. Con el objetivo de comprender mejor el comportamiento de las emergencias, se eligió realizar una simulación para obtener información sobre el tráfico, utilizando para ello la Routes API de Google Maps ⁴⁹ esta herramienta permite simular rutas aleatorias y recopilar datos cruciales para el estudio. Esta plataforma proporciona una gama de herramientas para desarrolladores y organizaciones interesadas en la geolocalización y la planificación de rutas. La solicitud API se diseñó para incluir coordenadas de origen y destino, recopilando información detallada sobre la ruta simulada. La configuración específica de estas solicitudes se detalla en el anexo 1 adjunto a este documento.

La respuesta obtenida de la API se presenta en formato JSON, destacando elementos clave como la distancia total de la ruta, la duración del recorrido y detalles específicos sobre diferentes segmentos de la ruta. Se ha adjuntado el anexo 2, que proporciona un ejemplo detallado de esta respuesta, ofreciendo una visión técnica de los datos recopilados mediante Google Maps Platform. Particularmente, se extrajeron datos relevantes, como la velocidad promedio de los vehículos en puntos específicos de la ruta. Además, se categorizó la fluidez del tráfico en distintos segmentos, proporcionando una representación de la situación del tráfico en el momento del recorrido. Estos detalles adicionales se integran directamente en los análisis subsiguientes, ofreciendo una comprensión más completa de las condiciones de tráfico a lo largo de las rutas simuladas. En la sección de metodologías se expondrá detalladamente el procedimiento seguido para lograr los datos mencionados.

5.3. PROCESAMIENTO DE DATOS

En el desarrollo de la investigación, la calidad y preparación de los datos desempeñaron un papel fundamental para garantizar una eficiencia válida del modelo. Por lo tanto, esta sección se utilizará para exponer las técnicas de manipulación, generación y compensación de los datos, además, sienta las bases para el análisis alrededor de los mismos.

El proceso del preprocesamiento de los datos no solo es un paso técnico, ya que si esta etapa no se efectúa correctamente, el análisis y otros procesos posteriores podrían verse afectados de manera significativa.

⁴⁹ Google Maps. *Routes API*. https://developers.google.com/maps/documentation/routes?hl=es_419. 2023.

Esta tarea representa un gran parte del proceso de integración de datos espaciales⁵⁰, temporales y datos de diversas fuentes, manipulando su complejidad y formato para la adaptación a vectores utilizados para la interpretación en el modelo de aprendizaje máquina.

5.3.1. Limpieza de datos y estandarización Como procedimiento inicial del tratamiento de datos, se procedió a unificar los archivos mensuales individuales proporcionados por el CRUE, para esta tarea se utilizó la herramienta de procesamiento de datos Tableau Prep Builder⁵¹. Posterior a la unificación se realizó a la estandarización de todos los formatos que se encontraban presentes en los conjuntos de datos CRUE y DTB, a su vez, se llevó a cabo la eliminación de campos innecesarios en el contexto de esta investigación.

5.3.2. Geocodificación La geocodificación es el proceso de asignar coordenadas geográficas (latitud y longitud) a una dirección postal o a cualquier otro tipo de ubicación geográfica. Es una técnica utilizada en los Sistemas de Información Geográfica (SIG) y en las Tecnologías de la Información Geográfica (TIG) para visualizar y analizar datos espaciales. La geocodificación es importante porque permite la integración de datos de diferentes fuentes y su representación en un mapa, lo que facilita la toma de decisiones basadas en la ubicación⁵².

Debido a la naturaleza de los datos obtenidos mediante el CRUE y la DTB, en los cuales su componente espacial está descrito mediante una dirección en lenguaje natural, la geocodificación es necesaria para conseguir variables geoespaciales precisas (longitud y latitud). A continuación, se describirá el proceso realizado para la geocodificación de los datos.

⁵⁰ Todas las operaciones espaciales y datos de georeferencia fueron ajustados al estándar EPSG 4686 MAGNA-SIRGAS

⁵¹ Tableau Software. *Tableau Prep Builder*. Ver. 2022.2.

⁵² Reinaldo Pérez Machado. *Proceso de geocodificación urbana: ejemplos de su aplicación en la ciudad de Barcelona*. Vol. XIII. Journal Abbreviation: [RCG] Revista Catalana de Geografia. Revista digital de geografia, cartografia i ciències de la Terra Publication Title: [RCG] Revista Catalana de Geografia. Revista digital de geografia, cartografia i ciències de la Terra. 29 de mayo de 2008. DOI: 10.13140/2.1.1858.5280.

Se utilizaron librerías como pandas⁴⁶, googlemaps⁵³ y gmaps⁵⁴, las cuales facilitaron la manipulación del conjunto de datos y la comunicación entre al ambiente local y la de la API de geocodificación de Google⁵⁵. Los pasos específicos seguidos para conseguir el resultado expuesto en la figura 10 son:

1. Importación de librerías.
2. Lectura y especificación del formato de los datos.
3. Utilización de la API de Google Maps para obtener información geográfica precisa (latitud y longitud) a partir de la dirección completa.
4. Almacenar la información geográfica.



Figura 10. Figura explicativa geocodificación. Creación propia.

5.3.3. Depuración espacial Debido a que la geocodificación es susceptible a errores de escritura o referencias coloquiales de lugares locales, se decidió llevar a cabo una depuración espacial utilizando el MGN de 2022 con el polígono del municipio de Bucaramanga³⁸. Esto permitió retener únicamente los datos que se encuentran dentro de dicho polígono, garantizando así una representación geográfica coherente para el análisis posterior. En la sección de metodología se profundiza en la distribución de

⁵³ Devangela. *googlemaps*. <https://pypi.org/project/googlemaps/>. Ver. 4.10.0. 26 de ene. de 2023.

⁵⁴ Swistakm. *gmaps*. <https://pypi.org/project/gmaps/>. Ver. latest.

⁵⁵ Google Maps. *Geocoding API*. <https://developers.google.com/maps/documentation/geocoding/overview?hl=es-419>. 2023.

estos datos y cómo afectó esta depuración al conjunto final.

Para llevar a cabo la depuración espacial, se utilizaron bibliotecas como pandas, geopandas⁵⁶, y folium⁵⁷. Esta última biblioteca facilita la representación visual de polígonos y puntos utilizando datos de OpenStreetMap⁴², un proyecto colaborativo para la creación libre de mapas e información geográfica. Los pasos seguidos incluyen la lectura de los datos posteriores a la geocodificación y la comparación espacial con el polígono del área metropolitana de Bucaramanga (véase la figura 11). Una vez realizada la comparación, se procedió a descartar todos los datos que no cumplieran con esta condición, almacenando únicamente aquellos que sí lo hacían.

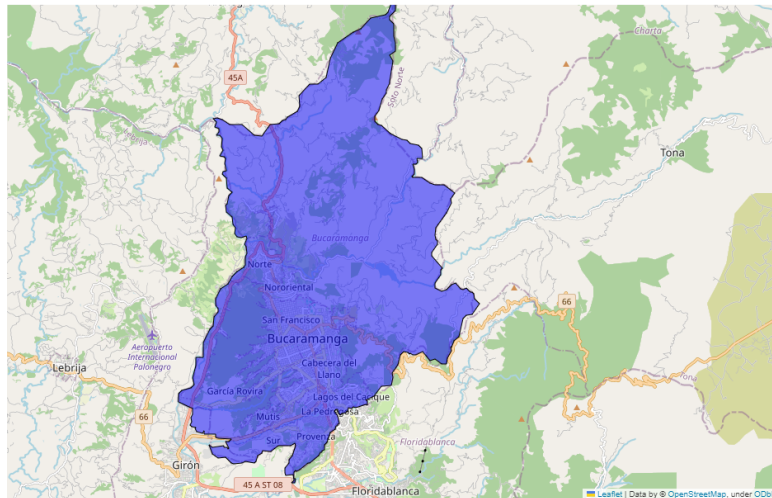


Figura 11. Polígono Area metropolitana de Bucaramanga. Creación propia con OpenStreetMap.

5.4. CONTRUCCIÓN DEL DATASET

En la preparación de los datos para la creación y evaluación de modelos predictivos resulta esencial una cuidadosa división del conjunto de datos en porciones de entrenamiento y prueba. Este proceso, fundamental para garantizar la validez y generalización del modelo, se llevó a cabo siguiendo una

⁵⁶ Kelsey Jordahl et al. *geopandas/geopandas: v0.8.1*. Ver. v0.8.1. Jul. de 2020. DOI: 10.5281/zenodo.3946761.

⁵⁷ python visualization. *Folium*. Ver. 0.11.0. 28 de dic. de 2020.

metodología específica, respaldada por el análisis previo de las series temporales de emergencias médicas y accidentes de tránsito.

5.4.1. Conjunto de Datos del CRUE Para el análisis de las emergencias médicas reportadas por el CRUE, se optó por dividir el conjunto de datos en periodos de entrenamiento y prueba (ver figura 12). En particular, se seleccionaron dos meses como datos de prueba, mientras que el resto de los datos fueron destinados al conjunto de entrenamiento. Los periodos específicos elegidos fueron:

- Conjunto de Entrenamiento (Entrenamiento): Enero de 2021 a octubre de 2022.
- Conjunto de Prueba (Prueba): Octubre de 2022 a enero de 2023.

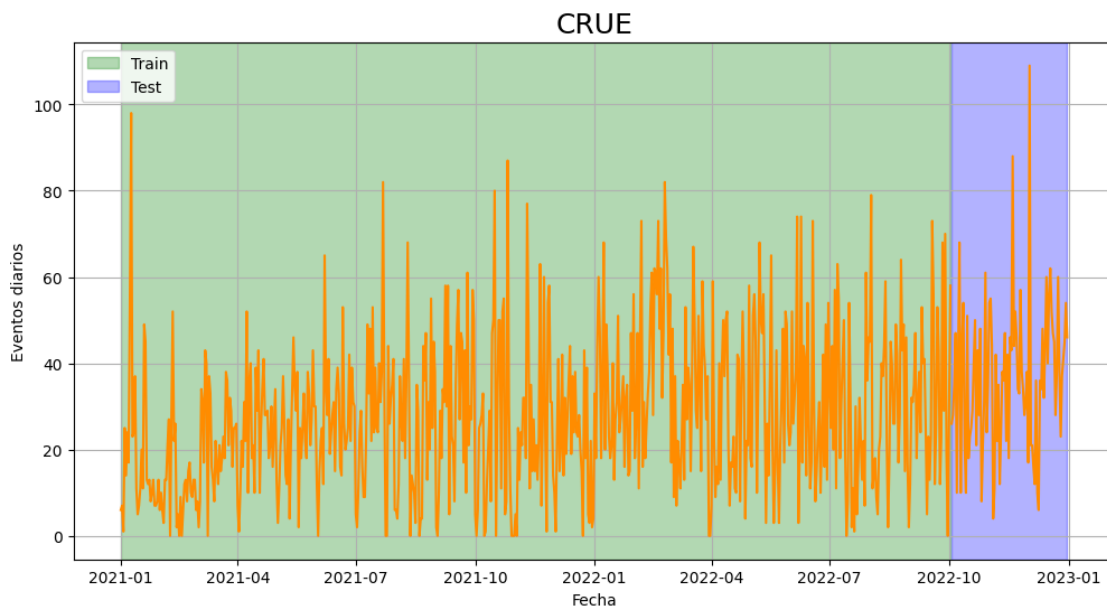


Figura 12. Selección de dataset CRUE. Creación propia.

Esta elección se basó en la necesidad de evaluar el rendimiento del modelo en un conjunto de datos independiente y representativo de situaciones no vistas durante la fase de entrenamiento.

5.4.2. Conjunto de Datos de la Dirección de Tránsito Debido a fenómenos como las restricciones de movilidad y cambios normativos, la construcción del conjunto de datos para la Dirección de Tránsito requirió una aproximación más detallada. Se crearon dos conjuntos de datos específicos:

1. Primer Conjunto de Datos (2017-2019):

- Entrenamiento: Enero de 2017 a diciembre de 2018.
- Prueba: Enero de 2019 a febrero de 2019.

2. Segundo Conjunto de Datos (2021-2022):

- Entrenamiento: Enero de 2021 a diciembre de 2021.
- Prueba: Enero de 2022 a febrero de 2022.

Esta segmentación se diseñó considerando fenómenos específicos en diferentes periodos, asegurando que el modelo se evaluara en situaciones relevantes y representativas (ver figura 13). La separación temporal de los datos permitirá la validación adecuada del modelo, considerando variaciones estacionales y cambios en el comportamiento del tránsito a lo largo del tiempo.

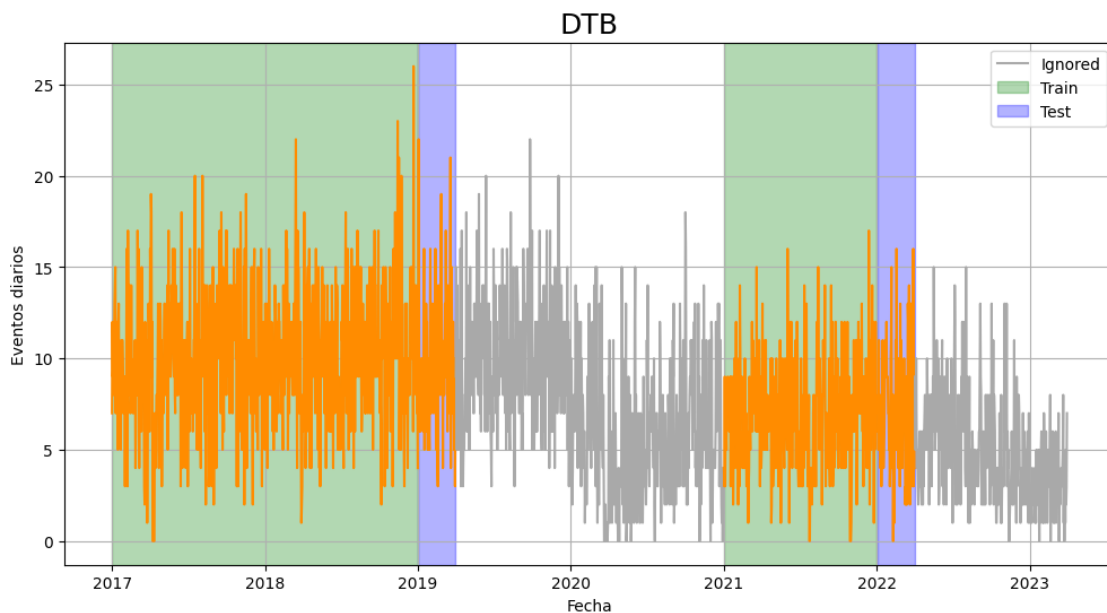


Figura 13. Selección de dataset DTB. Creación propia.

5.5. ANÁLISIS EXPLORATORIO DE LAS EMERGENCIAS MEDICAS

Una vez que se cuenta con la representación espacial y temporal de los datos, se procede al análisis descriptivo de las emergencias médicas y accidentes de tránsito. La selección de estos datos permite

asociar diversas variables, como las meteorológicas, temporales y viales, entre otras. Es importante destacar que este análisis no incluye los elementos detallados en los datos bases proporcionados por las entidades. En cambio, se centra únicamente en la frecuencia y comportamiento de los fenómenos durante los periodos analizados.

Para llevar a cabo un análisis de los datos, se procedió a la creación de gráficos representados en la figura 14 para los datos del CRUE y en la figura 15 para los datos DTB, estas figuras tiene una distribución similar en donde su parte superior representa una serie temporal con frecuencia diaria y la inferior una serie temporal con frecuencia mensual. Se utilizó la biblioteca matplotlib⁵⁸. Esta herramienta permite la creación de gráficos en dos dimensiones en Python. Además, se empleó statsmodels⁵⁹, una librería de exploración estadística de datos, para generar una gráfica embebida de tendencia en color naranja mediante el filtro de Hodrick-Prescott.

⁵⁸ J. D. Hunter. “Matplotlib: A 2D graphics environment”. En: *Computing in Science & Engineering* 9.3 (2007), págs. 90-95. DOI: 10.1109/MCSE.2007.55.

⁵⁹ Skipper Seabold y Josef Perktold. “statsmodels: Econometric and statistical modeling with python”. En: *9th Python in Science Conference*. 2010.

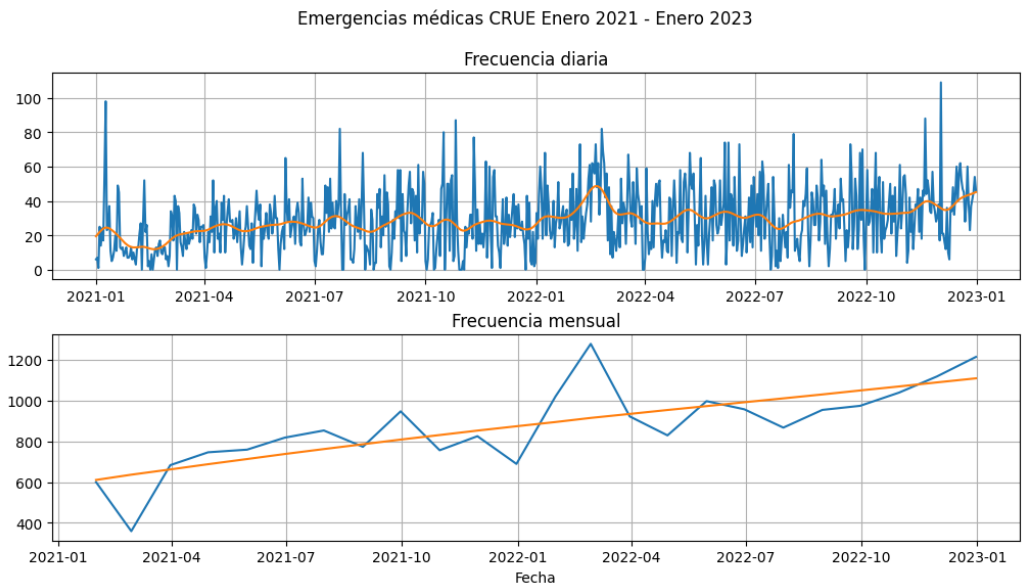


Figura 14. Emergencia médicas ocurridas en el área metropolitana de Bucaramanga. Creación por medio de la librería Matplotlib y análisis de tendencia con statsmodels.

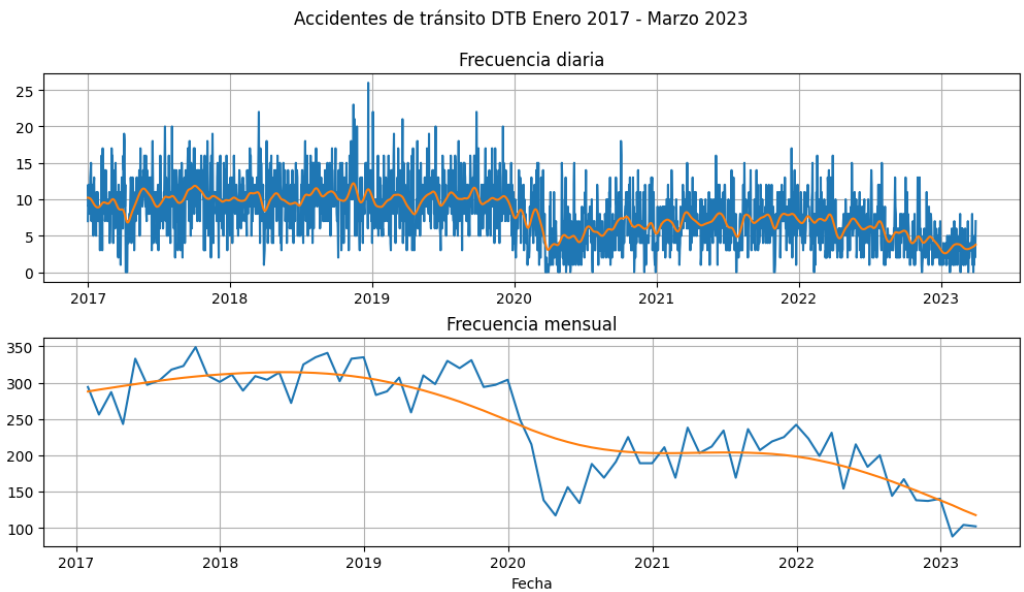


Figura 15. Accidentes de tránsito en el área metropolitana de Bucaramanga. Creación por medio de la librería Matplotlib y análisis de tendencia con statsmodels.

Según el autor Wolfgang Polasek ⁶⁰, el filtro de Hodrick-Prescott es un método de estimación paramétrica para obtener un componente de tendencia suave a través de la minimización de una función de pérdida para un parámetro de penalización λ fijo. El filtro de Hodrick-Prescott es popular en el análisis de ciclos económicos y ha sido utilizado por muchos investigadores para analizar puntos de inflexión en estos.

La fórmula general del filtro de Hodrick-Prescott 1 es la siguiente:

$$\min \sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2. \quad (1)$$

donde y_t es la serie de tiempo observada en el momento t , τ_t es el componente de tendencia suave en el momento t y λ es un parámetro de suavizado que controla la cantidad de suavizado en el componente de tendencia. La primera parte de la función de pérdida mide la precisión de ajuste del modelo y la segunda parte penaliza las fluctuaciones en el componente de tendencia.

En el caso de la figura 15, se puede observar un cambio significativo en el comportamiento de los datos en el año 2020, coincidiendo con los eventos asociados a la pandemia de COVID-19. Esto se debe a las restricciones a la movilidad implementadas por las autoridades locales con el objetivo de reducir los contagios. Asimismo, se aprecia una disminución en los registros a partir del primer semestre de 2022.

A raíz de esta observación, se indagó en las decisiones político-administrativas tomadas por el Congreso de Colombia o por las autoridades encargadas de la gestión del tránsito local. Como resultado, se encontró que la Ley 2161 de 2021⁶¹, modificada parcialmente por el Decreto 998 de 2022⁶², permite la retirada inmediata de los vehículos involucrados en accidentes sin lesiones personales. Esta medida acelera la liberación de la vía, reduce el impacto en el tráfico y mejora la fluidez.

⁶⁰ Wolfgang Polasek. *The Hodrick-Prescott (HP) filter as a Bayesian regression model*. Inf. téc. Reihe Ökonomie/Economics Series, 2011.

⁶¹ Congreso de Colombia. *Ley 2161 de 2021*. 26 de nov. de 2021.

⁶² Presidente de la república. *Decreto 998 de 2022*. 13 de jun. de 2022.

De acuerdo con los autores de esta ley, la rápida resolución de daños materiales mediante acuerdos entre las partes y las compañías aseguradoras podría resultar en una disminución en el número de casos intervenidos después de la entrada en vigencia de esta normativa, lo que implicaría una reducción en la documentación por parte de las autoridades de tránsito.

A partir de los comportamientos expuestos, se optó por seleccionar, en el caso de los datos de la dirección de tránsito, dos períodos de tiempo específicos para esta investigación. El primero abarcaba desde enero de 2017 hasta marzo de 2019, mientras que el segundo intervalo comprendía desde enero de 2021 hasta marzo de 2022. Esta elección se hizo con el objetivo de evitar que los datos se vean afectados por los fenómenos previamente descritos.

En relación con la distribución espacial de las emergencias médicas y los accidentes de tránsito, se utilizaron mapas de calor para visualizar la disposición geográfica de estos eventos en el área metropolitana de Bucaramanga. El Mapa de Calor de Emergencias Médicas (ver figura 16) resalta una concentración significativa en el centro de la ciudad, señalando áreas con una demanda potencialmente más alta de servicios de atención médica de emergencia. De manera similar, el Mapa de Calor de Accidentes de Tránsito (ver figura 17) revela una mayor incidencia en el mismo núcleo urbano, indicando áreas propensas a incidentes viales. Estos hallazgos subrayan la necesidad de medidas específicas para mejorar la seguridad vial y la atención de emergencias en las zonas identificadas como críticas en ambos mapas. La alta coincidencia de áreas críticas en ambos mapas resalta la importancia de un enfoque integrado en la planificación de intervenciones y la asignación de recursos.

Dada la complejidad y la naturaleza multifactorial de los datos geográficos y temporales asociados a las emergencias médicas y accidentes de tránsito en el área metropolitana de Bucaramanga, la elección de modelos de predicción adecuados es crucial. En este contexto, la elección de utilizar Random Forest se fundamenta en varias consideraciones.

Distribución geográfica de los accidentes de tránsito (DTB)

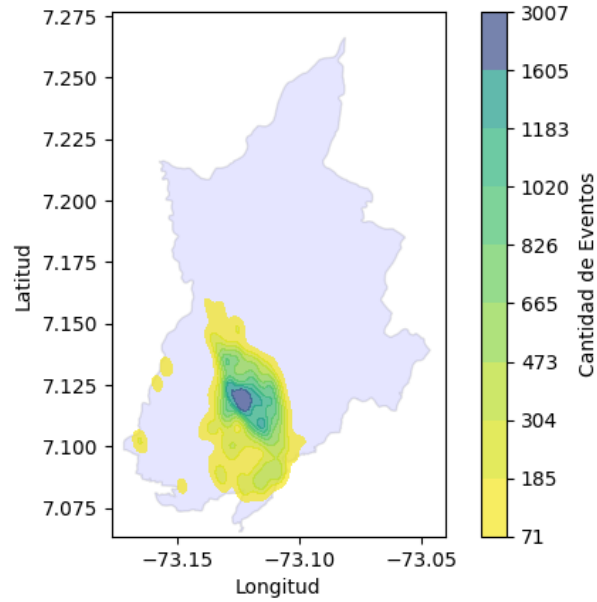


Figura 16. Distribución de la emergencia medicas en el área metropolitana de Bucaramanga. Creación propia con seaborn y matplotlib.

Distribución geográfica de las emergencias medicas (CRUE)

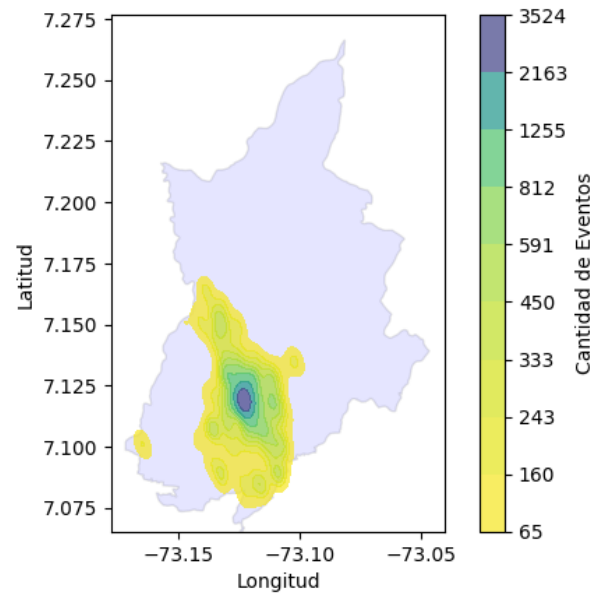


Figura 17. Distribución de accidentes de tránsito en el área metropolitana de Bucaramanga. Creación propia con seaborn y matplotlib.

6. METODOLOGÍA

6.1. REPRESENTACIÓN ESPACIAL

El enfoque seleccionado en esta investigación se centra en la representación espacial por intersecciones, como se menciona en ²⁷. Las intersecciones fueron elegidas debido a que son puntos de referencia geográficos claramente definidos y fácilmente identificables en la ciudad. Además, las intersecciones suelen ser lugares donde se concentra una considerable actividad humana y vehicular, lo que las convierte en sitios propensos a la ocurrencia de emergencias médicas.

Aunque existen otras representaciones espaciales, como cuadrículas o áreas irregulares, que podrían haber sido utilizadas en el estudio, presentan ciertas limitaciones. Por ejemplo, las cuadrículas pueden carecer de la precisión necesaria para capturar la variabilidad espacial de la ocurrencia de emergencias médicas en la ciudad, mientras que las áreas irregulares pueden resultar demasiado extensas para identificar patrones específicos del fenómeno de investigación.

Para iniciar este enfoque de representación, se procedió a recuperar las intersecciones. Se utilizó un modelo de representaciones espaciales que incluye las vías e intersecciones del área metropolitana de Bucaramanga. Este modelo contiene polígonos, polilíneas y puntos que representan las características principales de la malla vial. La biblioteca OSMnx⁴³ facilitó la recuperación de este objeto geoespacial. En la figura 18, se muestra una representación visual en la sección izquierda de las vías e intersecciones utilizando Networkx⁶³ y Matplotlib ⁵⁸.

Al recuperar inicialmente las intersecciones sin restricciones, se obtuvo un total de 4569 intersecciones, lo cual podría aumentar la complejidad al relacionar variables medioambientales, demográficas, viales, entre otras. Por lo tanto, fue necesario realizar una reducción en la cantidad de intersecciones mediante su simplificación. Este proceso se llevó a cabo mediante la *bufeización* de los nodos. La *bufeización*

⁶³ Aric Hagberg, Pieter Swart y Daniel S Chult. *Exploring network structure, dynamics, and function using NetworkX*. Inf. téc. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

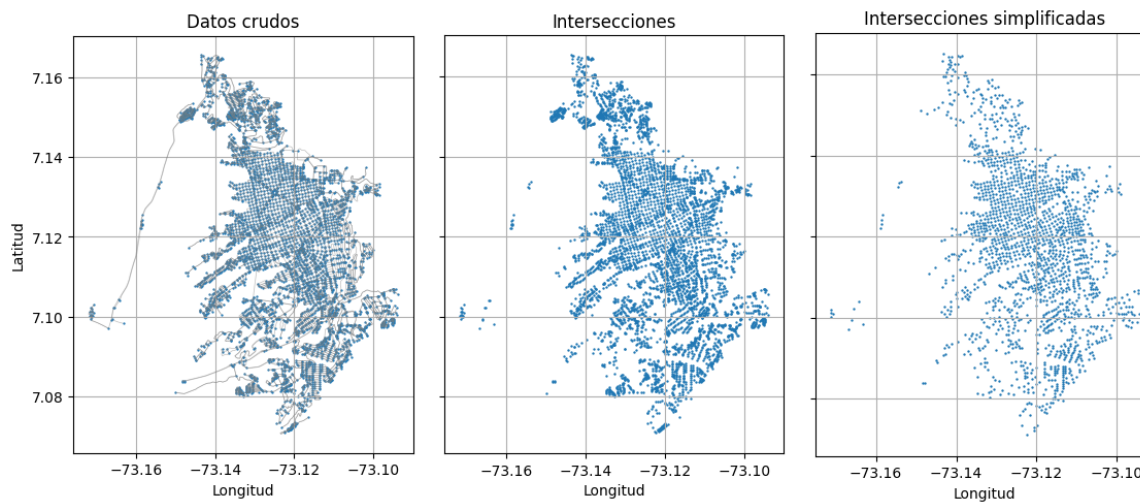


Figura 18. Comparación entre datos crudo, intersecciones originales y reducidas de Bucaramanga. Creación por medio de las librerías Matplotlib, OSMnx y Network.

según Guesgen⁶⁴ implica la creación de áreas circulares alrededor de cada intersección llamadas *buffers*, con el propósito de reducir la complejidad de la representación espacial, recuperando los *buffers* fusionados en un nuevo centroide único, esto verificando la topología de cada uno antes de fusionar. Esto evita que se fusionen nodos topológicamente remotos, pero espacialmente próximos. Este enfoque busca mantener una resolución interesante para la representación de emergencias médicas y accidentes de tránsito, asegurando al mismo tiempo que la información sea relevante y detallada. Es importante destacar que la *bufferización* contribuye a la simplificación de la red de calles al generar áreas que representan las intersecciones de manera más generalizada, lo cual puede ser beneficioso en situaciones específicas de análisis y visualización espacial.

En el caso de esta investigación se generaron *buffers* con un radio definido de 25 metros. La simplificación redujo la cantidad de intersecciones a 1904, permitiendo así mantener una resolución interesante en la representación espacial de las emergencias médicas y accidentes de tránsito.

En la figura 18 en la parte central y posteriormente la sección derecha, se visualiza la simplificación de

⁶⁴ Hans W. Guesgen et al. "Buffering Fuzzy Maps in GIS". En: *Spatial Cognition & Computation* 3.2 (1 de sep. de 2003). Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13875868.2003.9683762>, págs. 207-222. DOI: 10.1080/13875868.2003.9683762.

las intersecciones. Posteriormente, al relacionar estas intersecciones con los eventos correspondientes, se identificó que algunas intersecciones no tenían ningún evento asociado. Este proceso resultó en un conjunto de 1626 intersecciones que mantenían una relación con al menos un evento, proporcionando así una visión más específica y focalizada de las áreas de interés para el estudio.

6.2. REPRESENTACIONES TEMPORALES

La representación efectiva de datos cíclicos mediante técnicas de codificación y decodificación se ha convertido en un aspecto fundamental en diversas disciplinas. Este enfoque no solo permite una representación más compacta de los datos, especialmente en entornos donde el almacenamiento y la transmisión de datos son recursos limitados, sino que también desvela patrones y tendencias que podrían pasar desapercibidos en representaciones lineales convencionales. La capacidad de codificar y decodificar datos cíclicos se revela como una herramienta valiosa, no solo para analizar datos históricos, sino también para prever futuros valores en series temporales, desencadenando aplicaciones significativas en pronósticos climáticos, predicciones de precios de acciones y más.

En el contexto específico de la investigación en curso, la aplicación de técnicas de codificación y decodificación para representar datos temporales cíclicos revela su valía de manera palpable. La transformación de fechas en variables temporales, tales como día de la semana y mes, se convierte en un paso crucial para comprender la dinámica temporal de los eventos analizados⁶⁵.

Al emplear una representación trigonométrica para conferir un componente cíclico a estas variables temporales, se logra capturar la periodicidad inherente en fenómenos recurrentes. Esta codificación no solo facilita la identificación de patrones temporales complejos, sino que también aporta una dimensión adicional al análisis, permitiendo explorar la correlación entre eventos y ciclos temporales de manera más precisa.

La representación compacta de datos resultante de esta codificación se traduce en una mayor eficiencia en el procesamiento y análisis. La capacidad de discernir patrones y tendencias de manera más clara abre nuevas posibilidades para la formulación de hipótesis y la toma de decisiones informada.

⁶⁵ Anthony Adams y Peter Vamplew. "Encoding and Decoding Cyclic Data". En: *The South Pacific Journal of Natural Science* 16 (1 de ene. de 1998).

Así, en el marco de esta investigación, la utilización de técnicas avanzadas de codificación temporal no solo optimiza la manipulación de datos, sino que también enriquece la comprensión de la relación temporal entre eventos, aportando una perspectiva más completa a la exploración de emergencias médicas y accidentes de tránsito en el área metropolitana de Bucaramanga.

La aplicación concreta de la codificación temporal se llevó a cabo mediante la extracción de variables temporales relevantes de dos conjuntos de datos fundamentales: el Registro de Urgencias y Emergencias del CRUE y la Base de Datos de la DTB. Para llevar a cabo esta transformación, se optó por emplear la biblioteca Feature-engine, reconocida por su capacidad para realizar ingeniería de características de manera eficiente y precisa.

El proceso de transformación consistió en extraer información temporal clave, como el mes y el día de la semana, de los datos originales, posteriormente, se aplicó la codificación trigonométrica utilizando funciones seno y coseno para representar estos atributos temporalmente cíclicos. La biblioteca Feature-engine⁶⁶ facilitó este procedimiento, permitiendo una integración fluida y proporcionando herramientas especializadas para la manipulación de variables temporales.

El conjunto de gráficos expuestos en la figura 19, presenta una exploración detallada de la transformación de características temporales cíclicas, en este caso, asociadas a los meses del año. La representación en la sección izquierda destaca la ventana temporal de 48 meses frente a los meses del año, donde cada salto de mes está marcado de 1 a 12. Esta visualización revela la limitación inherente de este enfoque, ya que muestra discontinuidades anuales notables. Cuando el valor del mes pasa de 12 a 1, se produce un salto aparentemente significativo en el gráfico, evidenciando la dificultad en la interpretación temporal. A pesar de que la diferencia de tiempo entre registros adyacentes es de un mes, el salto del mes 12 al 1 crea una aparente discrepancia de -11. La codificación circular de características temporales, ilustrada en las secciones central y derecha de la figura mediante las transformaciones seno y coseno respectivamente, ofrece una alternativa más robusta. Ambas transformaciones presentan una diferencia absoluta constante de un mes antes y después de cada cambio, superando la limitación de la representación lineal y proporcionando una codificación más coherente de la información temporal asociada a los meses del

⁶⁶ Soledad Galli. “Feature-engine: A Python package for feature engineering for machine learning”. En: *Journal of Open Source Software* 6.65 (2021), pág. 3642. DOI: 10.21105/joss.03642.

año.

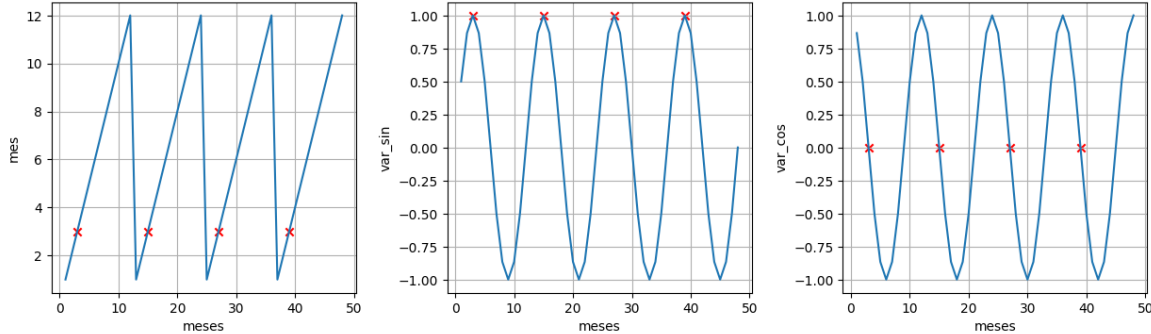


Figura 19. Representación de codificación mensual cíclica, ejemplificación del mes de marzo. Creación propia.

Como parte esencial de la comprensión de los gráficos, se adoptó la representación puntual del mes de marzo en la ventana de tiempo específica. Este mes se distingue por una marca "x" de color rojo en los gráficos. La representación experimenta variaciones en cada transformación, las cuales se llevan a cabo mediante las siguientes fórmulas de transformación circular:

$$var_sin = \sin \left(variable \left(\frac{2 \cdot \pi}{max_value} \right) \right). \quad (2)$$

$$var_cos = \cos \left(variable \left(\frac{2 \cdot \pi}{max_value} \right) \right) \quad (3)$$

Estas fórmulas 2 3 describen las transformaciones circulares mediante funciones seno y coseno, donde “variable” es la representación original del mes, semana o día, *max_value* indica el valor máximo que puede tomar la representación original. Estas transformaciones permiten visualizar la representación de los meses en diferentes contextos y resaltar patrones específicos en los datos.

Esta representación visual proporciona una visión intuitiva de la distribución y variación cíclica de las variables temporales, destacando la capacidad de la codificación trigonométrica para capturar la periodicidad inherente en datos temporales. La figura 19 se convierte así en una herramienta valiosa para validar la adecuada aplicación de la técnica de codificación y comprender su impacto en la estructura temporal de los conjuntos de datos analizados.

6.3. SIMULACIÓN DE TRÁFICO

La simulación de tráfico desempeña un papel crucial en la comprensión de la dinámica vehicular y la planificación eficiente de las vías urbanas. En esta sección, se presenta un enfoque innovador basado en la generación de rutas aleatorias y la obtención de información en tiempo real utilizando la API de Routes de Google Maps ⁴⁹. Este método, implementado a través de un script de Python diseñado para peticiones específicas, permite la creación y seguimiento de rutas simuladas en un entorno virtual, proporcionando valiosa información para la toma de decisiones en el ámbito de la seguridad vial y servicios de emergencia.

La generación de rutas se inicia a partir de un conjunto de intersecciones previamente simplificado y vinculado a eventos de emergencias médicas y accidentes de tránsito en el área metropolitana de Bucaramanga. Seleccionando dos intersecciones de este conjunto, se establecen como punto de inicio y destino respectivamente, creando así escenarios de viaje realistas y representativos de la complejidad del entorno urbano.

El período de simulación abarca un lapso de 20 días, desde el 8 de octubre de 2023 hasta el 25 de octubre de 2023. Durante este tiempo, el script realiza peticiones continuas a la API de Routes de Google Maps, generando rutas aleatorias y recopilando datos en tiempo real sobre el tráfico simulado. Este enfoque no solo permite evaluar la variabilidad de las condiciones de tráfico, sino que también facilita el análisis de posibles patrones asociados con eventos de emergencia, contribuyendo así a una comprensión más profunda de la interacción entre el tráfico vehicular y la ocurrencia de emergencias médicas y accidentes de tránsito.

En la fase inicial de la simulación de tráfico, se configuraron los parámetros esenciales para llevar a cabo el escenario de prueba. La simulación se diseñó para generar peticiones de rutas aleatorias utilizando la API Routes de Google. Se estableció un conjunto de intersecciones previamente simplificadas y relacionadas con emergencias médicas y accidentes, siendo este el punto de partida para la selección de nodos de inicio y fin en cada recorrido simulado.

La configuración básica de la simulación abordó la cantidad de peticiones a realizar por hora y definió los períodos durante los cuales se efectuarían estas solicitudes. Se fijó como base la generación de 100 peticiones por hora en el intervalo horario de 8 a. m. a 11 p. m. Si la simulación se encontraba fuera de

esta franja horaria, se reducía el número de peticiones a 20 por hora.

Dada la magnitud total de peticiones, que alcanzó las 27,808 solicitudes, se anticipó que el tratamiento de estos objetos JSON podría requerir una cantidad considerable de recursos y tiempo. En respuesta a esta consideración, se implementó una base de datos SQL para gestionar eficientemente estos objetos. Para la inserción de los datos de los objetos de respuesta JSON en la base de datos, se desarrolló un script en Python. El script, haciendo uso de bibliotecas como `os`, `json`, `pandas`, `mysql.connector` y `numpy`, facilitó la manipulación y extracción de la información deseada. Este enfoque permitió gestionar eficazmente la gran cantidad de datos generados durante la simulación, garantizando una operación más eficiente y un manejo adecuado de la información relevante para el análisis subsiguiente.

Con la base de datos implementada para la simulación de tráfico, se establecieron dos tablas fundamentales: *intersections* y *steps*. La tabla *intersections* (ver Figura 20) contiene dos atributos principales: *id* que identifica de manera única cada intersección y *coordinate* que almacena las coordenadas geográficas de la intersección en formato de punto. La clave primaria está definida por el campo *id*. Esta tabla actúa como un conjunto de intersecciones simplificadas relacionadas con eventos de emergencias médicas y accidentes de tráfico.

Por otro lado, la tabla *steps* (ver Figura 20) está diseñada para almacenar información detallada sobre los recorridos simulados. Sus campos incluyen *id* como clave única, *origin_id* y *destination_id* que se relacionan con las intersecciones de origen y destino respectivamente. El campo *coordinate* almacena las coordenadas del recorrido, *date* registra la fecha y hora de la simulación, *speed* representa la velocidad del recorrido, *velocity* la velocidad promedio, y *weekday* indica el día de la semana.

Ambas tablas están relacionadas mediante las claves foráneas *origin_id* y *destination_id* que apuntan a las intersecciones correspondientes en la tabla *intersections*.

Estos elementos estructurales forman la base de datos que respalda la simulación de tráfico, permitiendo un manejo eficiente de la información generada a partir de las peticiones realizadas a la API de Google Routes.

En una etapa posterior de la simulación de tráfico, se llevó a cabo un análisis detallado de los datos recopilados, buscando extraer información valiosa sobre el comportamiento del tráfico en el área metropolitana de Bucaramanga. Para lograr esto, se construyó una consulta SQL que agrupaba y resumía la información relevante, generando un compendio estructurado con los siguientes atributos:

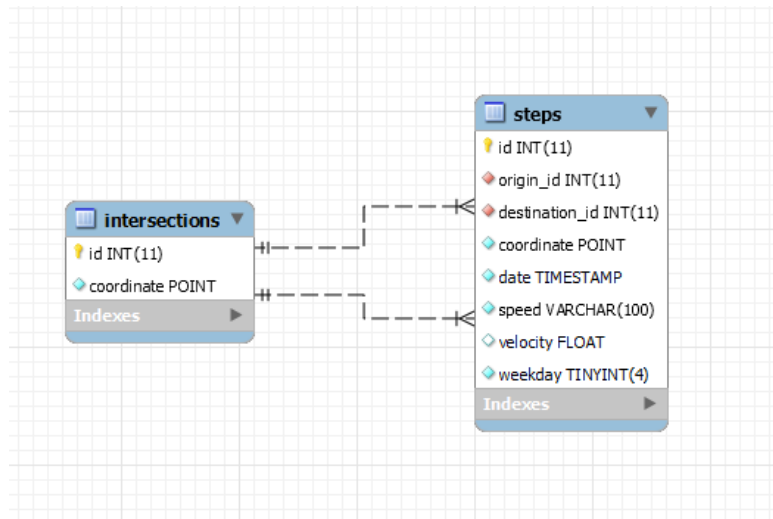


Figura 20. Diagrama relacional. Creación propia con MySQL workbench.

- Identificador único del registro.
- Identificación de la intersección.
- Día de la semana.
- Porcentaje de tiempo en estado NORMAL.
- Porcentaje de tiempo en estado SLOW.
- Porcentaje de tiempo en estado TRAFFIC_JAM.
- Velocidad promedio.

El compendio de datos proporciona una visión detallada de cómo se distribuyen los diferentes estados de tráfico en cada intersección durante los distintos días de la semana. Los porcentajes están determinados por la clasificación proporcionada por Google, reflejando la proporción de veces que una intersección estuvo en un estado específico en un día particular de la semana. Este compendio brinda una visión cuantitativa de la variabilidad del tráfico en diferentes momentos de la semana, permitiendo identificar patrones y tendencias que pueden ser cruciales para la planificación de la movilidad urbana y la respuesta a emergencias médicas y accidentes de tránsito. En la siguiente sección, se explorarán a fondo los resultados y las conclusiones derivadas de este análisis.

6.4. FORMACIÓN DE VECTORES DE CARACTERÍSTICAS

La metodología de ventana deslizante es una técnica ampliamente utilizada en el análisis y modelado de series temporales. Esta estrategia implica dividir la serie temporal en segmentos de longitud fija o variable, conocidos como ventanas, y realizar análisis o entrenamiento de modelos de manera secuencial en estos subconjuntos de datos⁶⁷.

En el contexto de la formación de vectores de características, integramos información de las emergencias médicas del CRUE y los accidentes de tránsito de las DTB. La primera etapa de este proceso implica la exploración de diversas configuraciones, considerando variables como la longitud de las ventanas y la frecuencia de deslizamiento. A continuación, analizamos el impacto en la capacidad predictiva de los modelos al incorporar estas características en los vectores de datos.

La optimización de la ventana deslizante con el objetivo de extraer la máxima información relevante para el modelo, mientras se mantiene una cantidad manejable de variables, tiene importantes implicaciones técnicas. En primer lugar, al determinar la longitud de las ventanas y la frecuencia de deslizamiento de manera eficiente, se logra un equilibrio crucial entre la capacidad del modelo para capturar patrones temporales y espaciales y la complejidad computacional asociada. Esto no solo favorece el rendimiento del modelo en términos de precisión, sino que también lo hace más eficiente en términos de recursos computacionales.

Además, la optimización de la ventana deslizante puede tener un impacto significativo en la viabilidad de incorporar modelos en sistemas embebidos o en entornos donde los recursos son limitados. Al reducir la dimensionalidad de los datos mediante la extracción de características clave, se facilita la integración de modelos en tiempo real y su despliegue en sistemas que requieren respuestas rápidas. Esta adaptabilidad es fundamental, por ejemplo, en el contexto de entornos urbanos inteligentes, donde la toma de decisiones en tiempo real basada en la predicción de emergencias médicas y accidentes de tránsito puede contribuir a una gestión más eficiente de los recursos y una respuesta más rápida ante

⁶⁷ Rey Francisco Ayala Castrejon y Christian Bucio Pacheco. “Modelo ARIMA aplicado al tipo de cambio peso-dólar en el periodo 2016-2017 mediante ventanas temporales deslizantes”. En: *Revista Mexicana de Economía y Finanzas Nueva Época REMEF* 15.3 (30 de jun. de 2020). Number: 3, págs. 331-354. DOI: 10.21919/remef.v15i3.466.

situaciones críticas. En última instancia, la optimización de la ventana deslizante no solo busca mejorar la capacidad predictiva de los modelos, sino también hacerlos más accesibles y aplicables en escenarios del mundo real.

La metodología de ventana deslizante no solo se limita a la segmentación de series temporales, sino que también se convierte en el punto focal donde aplicamos ingeniería de características. En este proceso, adaptamos y seleccionamos cuidadosamente las variables exógenas provenientes de las emergencias médicas del CRUE y los accidentes de tránsito de las DTB para formar vectores de características significativos. Este paso es crucial para mejorar la comprensión del modelo, ya que implica la identificación y extracción de las características más relevantes que contribuirán a la capacidad predictiva del mismo. Al emplear ingeniería de características en este contexto, no solo personalizamos los datos para satisfacer las necesidades específicas del modelo, sino que también simplificamos la interpretación de las variables exógenas, facilitando así la comprensión y el análisis del comportamiento de los eventos en el tiempo y el espacio.

6.5. MODELAMIENTO

El proceso de generación de predicciones se ha estructurado en un flujo metodológico que garantiza la coherencia y validez de los resultados. La figura 21 presenta una representación visual del flujo seguido para obtener predicciones a partir de los conjuntos de datos previamente preparados.

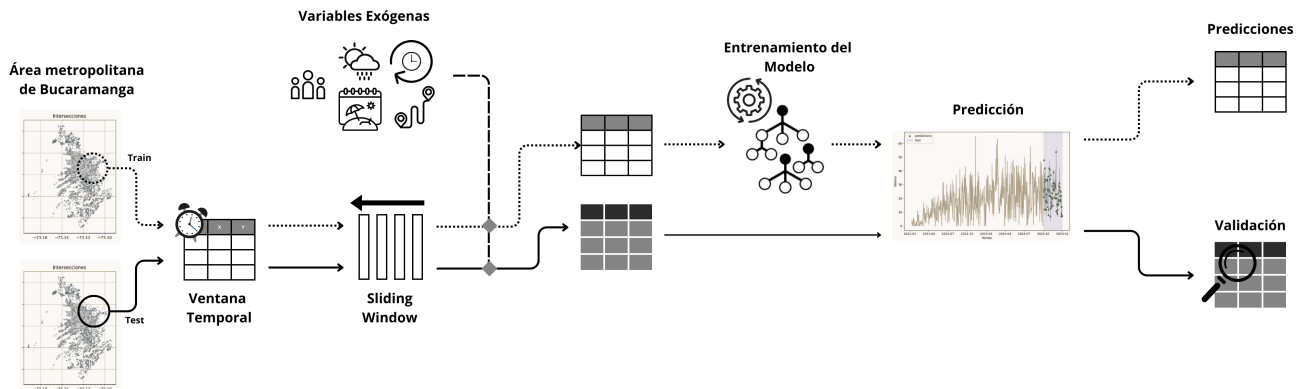


Figura 21. Diagrama de flujo. Creación propia.

El flujo consta de las siguientes etapas:

1. Representación de Intersecciones y Variables:

- Se inicia con la representación de las intersecciones seleccionadas en el área de estudio y las variables diversas expuestas a lo largo del documento. Estos conjuntos de datos deben relacionarse espacial o temporalmente para capturar las interacciones y patrones relevantes.

2. Aplicación de Ventana Deslizante:

- La frecuencia de los eventos, ya sean accidentes de tránsito o emergencias médicas, se somete a la técnica de ventana deslizante. Este enfoque permite dividir la serie temporal en segmentos de longitud fija, facilitando la captura de patrones temporales locales.

3. División de Datos para Entrenamiento y Evaluación:

- Como parte fundamental del tratamiento del modelo, los conjuntos de datos específicos se dividen en secciones de entrenamiento y evaluación. Esta división garantiza la adecuada validación del modelo y su capacidad para generalizar a datos no vistos durante el entrenamiento.

4. Creación de Vectores de Características:

- Los conjuntos de datos, ahora divididos y tratados, se utilizan para formar vectores de características que se introducirán en el modelo seleccionado. Estos vectores contienen información relevante y representativa para la generación de predicciones.

5. Entrenamiento del Modelo:

- Con los vectores de características listos, se procede al entrenamiento del modelo. Este paso implica el ajuste de los parámetros del modelo para que pueda aprender patrones y relaciones en los datos de entrenamiento.

6. Validación del Rendimiento del Modelo:

- Finalmente, el modelo entrenado se evalúa en los datos de evaluación para validar su rendimiento. Se analizan métricas específicas para determinar la capacidad predictiva del modelo y su capacidad para generalizar a situaciones no vistas durante el entrenamiento.

Este flujo metodológico garantiza un enfoque sistemático y robusto en la generación de predicciones, permitiendo obtener resultados confiables y útiles para la comprensión de la dinámica del tráfico y la gestión de emergencias médicas.

7. MODELO DE APRENDIZAJE AUTOMÁTICO

En este capítulo, se introduce el método de aprendizaje automático seleccionado para la construcción del modelo de predicción de emergencias médicas, así como los modelos complementarios para predecir accidentes de tránsito. Antes de profundizar en el método principal, es necesario definir el concepto de árboles de decisión, que constituyen una parte fundamental de este enfoque.

Los árboles de decisión desempeñan un papel crucial en el ámbito del aprendizaje automático y la inteligencia artificial, especialmente en contextos de regresión, al construir reglas efectivas para prever valores continuos a partir de ejemplos. Estos árboles ofrecen una representación transparente y comprensible de las reglas de decisión, lo que los convierte en herramientas valiosas en diversas aplicaciones del mundo real. La construcción de árboles de decisión basados en ejemplos ha demostrado ser robusta y ha encontrado éxito en la creación de sistemas basados en el conocimiento, subrayando así su relevancia en el campo del aprendizaje automático y la inteligencia artificial. Según J.R. Quinlan⁶⁸, un árbol de decisión se define como una estructura de datos que representa un conjunto de reglas de decisión. Cada nodo interno del árbol corresponde a una prueba en una variable de entrada, mientras que cada hoja representa una decisión o predicción continua. El proceso de construcción de un árbol de decisión implica seleccionar la variable de entrada más relevante en cada paso y dividir el conjunto de datos en subconjuntos más pequeños según los valores de esa variable. El objetivo principal es maximizar la pureza de los subconjuntos resultantes, asegurando que cada uno contenga principalmente ejemplos de una sola clase. El árbol resultante se puede utilizar para prever nuevos valores continuos en función de sus características de entrada.

Para ilustrar de manera práctica la estructura de composición de un árbol de decisión, se presenta la figura 22, que exhibe la representación de un ejemplo de árbol de decisión aplicado a la predicción de la cantidad de accidentes que ocurrirán al día siguiente, basándose en variables específicas. La figura

⁶⁸ J. R. Quinlan. "Induction of decision trees". En: *Machine Learning* 1.1 (1 de mar. de 1986), págs. 81-106. DOI: 10.1007/BF00116251.



Figura 22. Un árbol de decisión que muestra la cantidad de accidentes el día siguiente en una intersección particular. las cifras debajo de las hojas muestran la cantidad.

proporciona una visualización clara y detallada de cómo el árbol toma decisiones paso a paso, evaluando distintas características para prever la cantidad de accidentes. Este ejemplo concreto sirve como guía para comprender la lógica interna de un árbol de decisión y cómo se estructuran las reglas de predicción en función de las variables seleccionadas.

A partir de los árboles de decisión se forma el método utilizado en esta investigación llamado bosques aleatorios el cual consiste en la construcción de un conjunto o ensamble de árboles de decisión. En este enfoque, cada árbol se entrena de manera independiente utilizando un subconjunto aleatorio de las características disponibles en el conjunto de datos. Posteriormente, los resultados de todos los árboles se combinan para tomar una decisión final, ya sea en el contexto de clasificación o regresión. Esta estrategia de ensamble y la introducción de aleatoriedad en la construcción de los árboles individuales son aspectos fundamentales de los bosques aleatorios.

Según el enfoque propuesto por Leo Breiman ¹⁸, los bosques aleatorios representan una extensión de los árboles de decisión, utilizando múltiples árboles para realizar predicciones. Cada árbol en el bosque se entrena de forma independiente mediante una muestra aleatoria del conjunto de datos, y la predicción final se obtiene al promediar las predicciones de todos los árboles individuales. Esta técnica se

basa en la premisa de que la combinación de varios modelos puede mejorar la precisión y robustez en comparación con un solo árbol de decisión. Los bosques aleatorios han demostrado su efectividad en diversas aplicaciones, incluyendo clasificación y regresión, siendo una extensión relevante del enfoque de los árboles de decisión en el ámbito del aprendizaje automático.

Un factor importante mencionado por Leo Breiman ¹⁸ es la selección precisa de hiperparámetros juega un papel crucial en la optimización del rendimiento del modelo de bosques aleatorios. La configuración óptima de estos hiperparámetros no solo influye en la calidad de las predicciones, sino que también desempeña un papel clave en la capacidad del modelo para adaptarse a nuevos datos sin caer en problemas como el sobreajuste.

Leo Breiman ¹⁸ sugiere que, al afinar cuidadosamente los hiperparámetros, se puede lograr un equilibrio adecuado entre la capacidad del modelo para ajustarse a los datos de entrenamiento y su habilidad para generalizar a situaciones no vistas. Este proceso de ajuste de hiperparámetros es esencial para evitar la sobreoptimización del modelo a los datos de entrenamiento específicos y, en cambio, fomentar una capacidad más robusta de adaptación a diversas condiciones. Siguiendo las perspectivas generales de los bosques aleatorio es pertinente la exploración de las virtudes frente al fenómeno estudiado.

En la investigación .Aplicación de métodos de aprendizaje automático a la interpolación espacial de variables ambientales realizada por Li Jin⁶⁹, se evaluó la eficacia de técnicas de aprendizaje automático, como Bosques Aleatorios (RF) y Máquinas de Soporte Vectorial (SVM), en la interpolación espacial de variables ambientales. El enfoque central fue comparar el rendimiento de 23 métodos distintos, que incluían RF, SVM, Kriging Ordinario (OK), Inverse Distance Squared (IDS) y sus combinaciones, para predecir la concentración de lodo en el margen suroeste de Australia. Los resultados revelaron que RF, especialmente al combinarse con OK o IDS, superó a los otros métodos en términos de precisión predictiva. Se resaltó la menor sensibilidad de RF a variables no importantes y ruidosas, contribuyendo a su sólido rendimiento incluso en presencia de tales variables.

⁶⁹ Jin Li et al. "Application of machine learning methods to spatial interpolation of environmental variables". En: *Environmental Modelling & Software* 26.12 (dic. de 2011), págs. 1647-1659. DOI: 10.1016/j.envsoft.2011.07.004.

Entre las ventajas asociadas al uso de RF se encuentran:

1. Capacidad para modelar interacciones complejas entre variables predictoras
2. Relativa robustez ante valores atípicos y ruido.
3. Selección implícita de variables importantes, reduciendo la sensibilidad a variables irrelevantes.
4. Superioridad en la gestión de datos poco predecibles en comparación con árboles de decisión, redes neuronales y máquinas de soporte vectorial.
5. Árboles individuales no podados, permitiendo obtener árboles con bajo sesgo y un buen rendimiento predictivo, incluso en presencia de variables predictoras con ruido.
6. Menor propensión al sobreajuste gracias a la aplicación de la ley de los grandes números.

8. EXPERIMENTOS Y RESULTADOS

En este capítulo, como primera parte, se procede a la descripción de las métricas utilizadas para evaluar la evolución del modelo y su configuración inicial, junto con el método para explorar los hiperparámetros. Se lleva a cabo la ejecución de experimentos para evaluar la efectividad de los modelos de predicción desarrollados y se realiza un análisis de los resultados obtenidos. Se presenta una visión detallada de la metodología experimental, los conjuntos de datos utilizados y los modelos implementados.

8.1. CONFIGURACIÓN EXPERIMENTAL

8.1.1. Métricas de evaluación La evaluación del rendimiento de un modelo es esencial para comprender su capacidad predictiva y su capacidad para generalizar a datos no vistos durante el entrenamiento. En este contexto, se utilizan dos métricas comunes: R^2 (Coeficiente de Determinación) y ECM (Error Cuadrático Medio). A continuación, se proporciona una descripción de cada métrica y su utilidad en la evaluación de modelos.

Coeficiente de Determinación (R^2) El coeficiente de determinación, comúnmente conocido como R^2 , es una métrica que evalúa la proporción de la variabilidad en la variable dependiente que es explicada por el modelo. R^2 tiene valores en el rango de $-\infty$ a 1, donde:

- $R^2 = 1$ El modelo explica toda la variabilidad de la variable dependiente.

La fórmula 4 para R^2 se expresa como:

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2} = 1 - \frac{\sigma_r^2}{\sigma_y^2} \quad (4)$$

donde σ_r^2 es la suma de los cuadrados de los residuos y σ_y^2 es la suma total de los cuadrados. Un valor cercano a 1 indica un buen ajuste del modelo a los datos ⁷⁰.

⁷⁰ Davide Chicco, Matthijs J. Warrens y Giuseppe Jurman. “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation”. En: *PeerJ. Computer Science* 7 (2021), e623. DOI: 10.7717/peerj-cs.623.

Error Cuadrático Medio (ECM) El Error Cuadrático Medio, o ECM por sus siglas en inglés, mide la magnitud promedio de los errores entre las predicciones del modelo y los valores reales. Es particularmente útil para evaluar la precisión del modelo en términos de la magnitud de los errores, penalizando fuertemente los errores grandes.

La fórmula 5 para ECM se expresa como:

$$ECM = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (5)$$

(mejor valor = 0; peor valor = $+\infty$)

donde n es el número de observaciones, Y_i son los valores reales, y \hat{Y}_i son las predicciones del modelo⁷¹.

8.1.2. Configuración de los vectores de características En esta sección, se detalla la configuración experimental que se llevará a cabo, incluyendo el conjunto de experimentos planificados. La configuración experimental es esencial para garantizar la reproducibilidad y comprensión de los resultados obtenidos.

Una vez completado el proceso de preprocesamiento de datos, que incluyó la geocodificación y el relacionamiento con las intersecciones seleccionadas del área de estudio, el enfoque principal se centra en la construcción de la ventana deslizante. Dada la granularidad diaria de la presente investigación, las configuraciones de ventana deslizante se han diseñado para cumplir con esta frecuencia. Las configuraciones de ventana deslizante planificadas son las siguientes:

1. Ventana de 7 días.
2. Ventana de 14 días.
3. Ventana de 21 días.
4. Ventana de 28 días.

A partir de estas, se introduce una variable que representa el conteo total de eventos en diferentes

⁷¹ “Mean Squared Error”. En: *The Concise Encyclopedia of Statistics*. New York, NY: Springer, 2008, págs. 337-339. DOI: 10.1007/978-0-387-32833-1_251.

periodos, configuradas de la siguiente manera:

- Ventana de 7 días y una variable total de 14 días.
- Ventana de 7 días y una variable total de 21 días.
- Ventana de 7 días y una variable total de 28 días.
- Ventana de 14 días y una variable total de 21 días.
- Ventana de 14 días y una variable total de 28 días.
- Ventana de 21 días y una variable total de 28 días.

Posteriormente, se recuperan varios conjuntos de datos que complementan la información temporal, incluyendo representaciones de días festivos, variables climatológicas, variables producto de la simulación de tráfico, descripciones viales a través de OSMnx y datos demográficos. Estos conjuntos de datos se mezclan con los datos primarios siguiendo procesos específicos. Además, lleva a cabo la ingeniería de características, generando nuevas variables temporales, tales como el día del mes, día en el año, día de la semana, semana del año, semana del mes y mes. Al igual que las configuraciones mencionadas, estas variables se someten a experimentación para evaluar su contribución a los modelos de predicción. Se exploran cambios en la evaluación del modelo al aplicar representaciones cíclicas a estas variables, tal como se describe en secciones anteriores.

Dado que los conjuntos de datos utilizados en la investigación presentan magnitudes diferentes, se hace necesaria la aplicación de un proceso de normalización. Este procedimiento tiene como objetivo ajustar las escalas de las variables para que puedan ser comparadas y utilizadas de manera más efectiva en los modelos de predicción.

Se implementaron tres métodos de normalización comunes:

1. **Min-Max Scaling:** Escala los datos al intervalo $[0, 1]$. Este método es especialmente útil cuando se quiere preservar la relación proporcional entre las variables 6.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (6)$$

Donde: X' = Valor normalizado

X = Valor original

X_{min} = Valor mínimo de los datos

X_{max} = Valor máximo de los datos

2. **Z-score Normalization:** Transforma los datos para que tengan una media de 0 y una desviación estándar de 1. Es útil cuando se asume que los datos tienen una distribución normal 7.

$$X' = \frac{X - \mu}{\sigma} \quad (7)$$

Donde: X' = Valor normalizado

X = Valor original

μ = Media de los valores

σ = Desviación estándar de los datos

3. **Robust Scaling:** Utiliza la mediana y el rango intercuartílico para escalar los datos, siendo menos sensible a valores atípicos que Min-Max Scaling. Es preferible cuando los datos contienen outliers 8.

$$X' = \frac{X - \mu}{IQR} \quad (8)$$

Donde: X' = Valor normalizado

X = Valor original

μ = Media de los valores

IQR = Rango intercuartil (Q3 - Q1)

La implementación de Random Forest en este estudio se llevó a cabo utilizando la biblioteca scikit-learn (sklearn), una herramienta eficiente y versátil para el aprendizaje automático en Python. scikit-learn proporciona una interfaz fácil de usar para varios algoritmos, incluido Random Forest, lo que simplifica la experimentación y ajuste de parámetros.

Al implementar Random Forest con scikit-learn, se tienen acceso a varios atributos que permiten un

mayor control y comprensión del modelo. Algunos de los atributos clave incluyen:

1. `n_estimators`: Número de árboles en el bosque.
2. `max_depth`: Profundidad máxima de los árboles.
3. `min_samples_split`: Número mínimo de muestras requeridas para dividir un nodo interno.
4. `min_samples_leaf`: Número mínimo de muestras requeridas en una hoja.
5. `max_features`: Número máximo de características a considerar para dividir un nodo.

Estos atributos son esenciales para ajustar y optimizar el rendimiento del modelo durante la experimentación. Se explorará cómo variar estos parámetros afecta a la capacidad predictiva del modelo en el contexto de la predicción de eventos médicos de emergencia y accidentes de tránsito.

Antes de realizar la búsqueda de los mejores hiperparámetros, se empleó la configuración por defecto proporcionada por la biblioteca `scikit-learn` para la implementación de Random Forest. La configuración por defecto incluye valores predeterminados para atributos clave que definen el comportamiento del modelo. Algunos de estos valores por defecto son:

1. `n_estimators`: 100 (Número de árboles en el bosque).
2. `max_depth`: None (Profundidad máxima de los árboles, sin límite por defecto).
3. `min_samples_split`: 2 (Número mínimo de muestras requeridas para dividir un nodo interno).
4. `min_samples_leaf`: 1 (Número mínimo de muestras requeridas en una hoja).
5. `max_features`: 'auto' (Número máximo de características a considerar para dividir un nodo).

Estos valores son utilizados como punto de partida antes de realizar la búsqueda de los mejores hiperparámetros. La experimentación posterior incluirá ajustes en estos valores para evaluar su impacto en la capacidad predictiva del modelo.

8.2. EXPERIMENTOS

En esta sección, se describen en detalle los experimentos realizados, abordando la configuración experimental, los conjuntos de datos utilizados 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.

8.2.1. Experimentos con el Dataset CRUE En la tabla 3, se presenta una codificación de algunos de los experimentos más representativos realizados con el dataset del CRUE. Sin embargo, es importante destacar que este trabajo incluyó una exhaustiva evaluación que abarcó más de 40 experimentos solo en el dataset del CRUE. Estos experimentos abordaron diversas configuraciones de ventanas deslizantes, pruebas con diferentes métodos de normalización, aplicaciones a cada uno de los conjuntos de datos y un examen detallado de las variables resultantes de la ingeniería de características. La selección de los experimentos presentados en la tabla se basa en su relevancia y aporte significativo al rendimiento general del modelo.

Código	Descripción del Experimento
E1	Evaluación del modelo con ventana deslizante de 7 días y variable acumulativa de 14 días.
E2	Evaluación del modelo con ventana deslizante de 7 días, variable acumulativa de 14 días y día del mes.
E3	Evaluación del modelo con ventana deslizante de 7 días, variable acumulativa de 14 días, día del mes y semana del mes.
E4	Evaluación del modelo con ventana deslizante de 7 días, variable acumulativa de 14 días, día del mes y semana del mes en representación circular.
E5	Evaluación del modelo con aplicación de normalización (Min-Max scaling)
E6	Evaluación del modelo porcentaje de frecuencia de trayecto TRAFIC_JAM

Tabla 3. Codificación de Experimentos CRUE

8.2.2. Experimentos con el Dataset DTB De manera similar, se realizaron experimentos exhaustivos con el dataset de la DTB. En las tablas 4 y 5, se presenta una codificación de algunos de los experimentos más representativos. Al igual que con el dataset del CRUE, se llevaron a cabo diversos experimentos para explorar las configuraciones y evaluar el rendimiento del modelo de manera integral.

Código	Descripción del Experimento
E7	Evaluación del modelo con ventana deslizante de 28 días.
E8	Aplicación de normalización (Min-Max scaling)
E9	Evaluación del modelo con velocidad promedio en la intersección.
E10	Evaluación del modelo con días festivos.
E11	Evaluación del modelo con días festivos laborales (celebraciones sociales).

Tabla 4. Codificación de Experimentos DTB (2017-2019)

Código	Descripción del Experimento
E12	Evaluación del modelo con ventana deslizante de 14 días y dato acumulativo de 21 días.
E13	Aplicación de normalización (Min-Max scaling)
E14	Evaluación del modelo con días festivos.
E15	Evaluación del modelo con días festivos y días laborales.
E16	Evaluación del modelo con semana del año.
E17	Evaluación del modelo con velocidad promedio en la intersección.

Tabla 5. Codificación de Experimentos DTB (2021-2022)

8.3. RESULTADOS

En esta sección, se presentan los resultados obtenidos de los experimentos. Se analizan las métricas de evaluación, como R^2 y MSE, para comprender el rendimiento de los modelos en diferentes configuraciones.

8.3.1. Optimización de Hiperparámetros La optimización de hiperparámetros es un paso crucial en el desarrollo de modelos de aprendizaje automático, ya que determina la configuración óptima que maximiza el rendimiento del modelo en datos no vistos. En esta sección, abordamos el proceso de optimización de hiperparámetros y cómo se aplica en el modelo.

Inicialmente, se lleva a cabo la evaluación del modelo utilizando su configuración base, que es la configuración predeterminada sin ajustes específicos. Esto proporciona una referencia inicial para comprender el rendimiento del modelo sin ninguna optimización de hiperparámetros.

Una vez completada la evaluación con la configuración base, se procederá a la optimización de hiperparámetros. Para ello, se utilizará la técnica de búsqueda aleatoria. En este enfoque, se seleccionan combinaciones aleatorias de hiperparámetros del espacio de búsqueda definido, y se evalúa el rendimiento del modelo para cada combinación.

El objetivo principal de la optimización es maximizar el coeficiente de determinación R^2 , ya que es una métrica fundamental para medir la calidad del ajuste del modelo. La búsqueda aleatoria se llevará a cabo de manera iterativa, explorando diferentes conjuntos de hiperparámetros para identificar la configuración que conduzca al mejor rendimiento en términos de R^2 .

Una vez realizada la optimización de los hiperparámetros a cada uno de los modelos, podemos observar los resultados obtenidos a continuación:

Parámetro	Valor Optimizado
n_estimators	124
bootstrap	True
max_depth	7
max_features	5
min_samples_leaf	9
min_samples_split	7

Tabla 6. Resultados de la optimización de hiperparámetros para el modelo CRUE.

Parámetro	Valor Optimizado
n_estimators	54
bootstrap	False
max_depth	12
max_features	10
min_samples_leaf	6
min_samples_split	5

Tabla 7. Resultados de la optimización de hiperparámetros para el modelo STB (2017-2019).

Parámetro	Valor Optimizado
n_estimators	186
bootstrap	False
max_depth	2
max_features	13
min_samples_leaf	9
min_samples_split	7

Tabla 8. Resultados de la optimización de hiperparámetros para el modelo DTB (2021-2022).

Los resultados después de la optimización de los hiperparámetros nos indican que: el modelo específico del CRUE obtuvo un valor de MSE de 0.005 y un valor R^2 de 0.351. Los modelos de la DTB obtuvieron un valor de MSE 0.018 y R^2 de 0.150 para el periodo de (2017-2019) y para el periodo (2021-2022) MES de 0.006 y R^2 de 0.177.

8.3.2. Resultados de Experimentos A continuación se muestran los resultados de los experimentos para cada modelo:

Configuración	R^2	MSE
E1	0.283	1.734
E2	0.314	1.659
E3	0.316	1.654
E4	0.323	1.639
E5	0.315	0.006
E6	0.334	0.006

Tabla 9. Resultados de los experimentos para el modelo CRUE.

Los resultados de los experimentos para el modelo CRUE se resumen en la tabla 9, que presenta los valores de las métricas mencionadas para cada configuración (E1 a E6). El coeficiente de determinación varía entre 0.283 y 0.334, siendo el valor más alto asociado a la configuración E6, indicando un ajuste más efectivo del modelo. En cuanto al MSE, que cuantifica la magnitud promedio de los errores entre las predicciones y los valores reales, se observa una variación entre 0.006 y 1.734. Los valores más bajos de MSE se registran en las configuraciones E5 y E6, destacando la configuración E6 con el coeficiente de determinación más alto y el MSE más bajo. Estos resultados sugieren que la configuración E6 ofrece un equilibrio óptimo entre ajuste del modelo y precisión en las predicciones para el modelo CRUE.

Configuración	R^2	MSE
E7	0.095	0.078
E8	0.091	0.020
E9	0.111	0.077
E10	0.112	0.077
E11	0.118	0.019

Tabla 10. Resultados de los experimentos para el modelo DTB (2017-2019).

La tabla 10 proporciona una visión detallada de los resultados obtenidos a partir de diferentes configuraciones experimentales aplicadas al modelo DTB (2017-2019). Cada fila representa una configuración específica, identificada por el código E7, E8, E9, E10 y E11, y se evalúan dos métricas clave descritas

anteriormente. El coeficiente de determinación varía entre 0.091 y 0.118, siendo el valor más alto asociado a la configuración E11, lo cual indica un ajuste más efectivo del modelo. En cuanto al MSE, que mide la magnitud promedio de los errores entre las predicciones y los valores reales, se observa que los valores oscilan entre 0.019 y 0.078, siendo el más bajo correspondiente a la configuración E11. Estos resultados señalan que la configuración E11 logra un equilibrio destacado entre un ajuste efectivo del modelo y una mayor precisión en las predicciones para el modelo DTB (2017-2019).

Configuración	R^2	MSE
E12	0.123	0.221
E13	0.115	0.006
E14	0.117	0.006
E15	0.118	0.006
E16	0.128	0.006
E17	0.130	0.006

Tabla 11. Resultados de los experimentos para el modelo DTB (2021-2022).

Los resultados de los experimentos para el modelo DTB en el periodo 2021-2022 se resumen en la tabla 11. El coeficiente de determinación muestra una variación entre 0.115 y 0.130 para las configuraciones E13 a E17. El valor más alto de se asocia con la configuración E17, indicando un ajuste más efectivo del modelo para explicar la variabilidad en los datos observados durante este periodo.

Al considerar el Error Cuadrático Medio (MSE), que cuantifica la magnitud promedio de los errores entre las predicciones del modelo y los valores reales, se observa una consistencia en los valores más bajos. Las configuraciones E13, E14, E15, E16 y E17 comparten un MSE de 0.006, destacando una coherencia en la capacidad de estas configuraciones para minimizar la discrepancia entre las predicciones y los resultados reales.

8.4. VISUALIZACIÓN DE RESULTADOS

La visualización de resultados es esencial para la comprensión y comunicación efectiva de los hallazgos. En esta sección, se incluyen gráficos de representaciones visuales relevantes. Estas visualizaciones

ayudan a contextualizar la información y proporcionan una perspectiva intuitiva sobre la distribución espacial y temporal de las emergencias médicas y accidentes de tránsito en el área de estudio.

El gráfico 23 de dispersión generado ofrece una representación visual de la relación entre las observaciones reales y las predicciones del modelo para el conjunto de prueba del CRUE. Cada punto en el gráfico representa una instancia única, donde el eje horizontal refleja las observaciones reales y el eje vertical representa las predicciones correspondientes del modelo. Idealmente, los puntos deberían distribuirse a lo largo de una línea diagonal, indicando una correspondencia perfecta entre las predicciones y los valores reales. Sin embargo, se observan desviaciones de esta línea ideal, lo que sugiere variaciones en la precisión de las predicciones en diferentes situaciones. Este análisis visual es esencial para evaluar la capacidad del modelo para generalizar y capturar patrones en los datos de manera coherente. Cualquier dispersión significativa podría indicar áreas de mejora en el rendimiento del modelo.

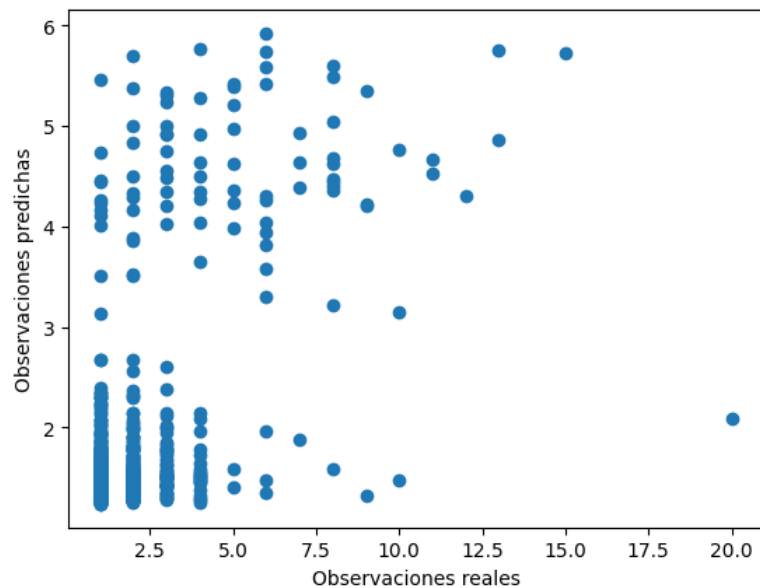


Figura 23. Gráfico de dispersión que compara las observaciones reales y las observaciones predichas por el modelo CRUE. Ambos conjuntos de datos están normalizados utilizando la escala min-max. Los resultados obtenidos se refleja en las bajas métricas de Error Cuadrático Medio (MSE) de 0.006 y un coeficiente de determinación de 0.334. Estos valores indican una precisión notable y un ajuste efectivo del modelo para explicar la variabilidad en los datos durante el período analizado

El gráfico de importancia de características del modelo del CRUE ver figura 24, basado en el algoritmo de Bosques Aleatorios, ofrece una valiosa perspectiva sobre las variables que más impactan en la capacidad predictiva del modelo. Al analizar el gráfico, se observa que las características más influyentes incluyen la variable acumulativa de eventos en los últimos 14 días, seguida de cerca por la cantidad de eventos registrados el quito día en la ventana deslizante y el séptimo día en la ventana deslizante. Estas variables emergen como factores cruciales que inciden significativamente en las predicciones del modelo. La capacidad de anticipar la ocurrencia de emergencias médicas se apoya fuertemente en la dinámica temporal acumulativa de los eventos y en los patrones diarios específicos. Este análisis contribuye de manera esencial a la interpretación del modelo y proporciona perspectivas valiosas para futuras iteraciones y mejoras, consolidando la fiabilidad de las predicciones en el ámbito de la asistencia médica de emergencia.

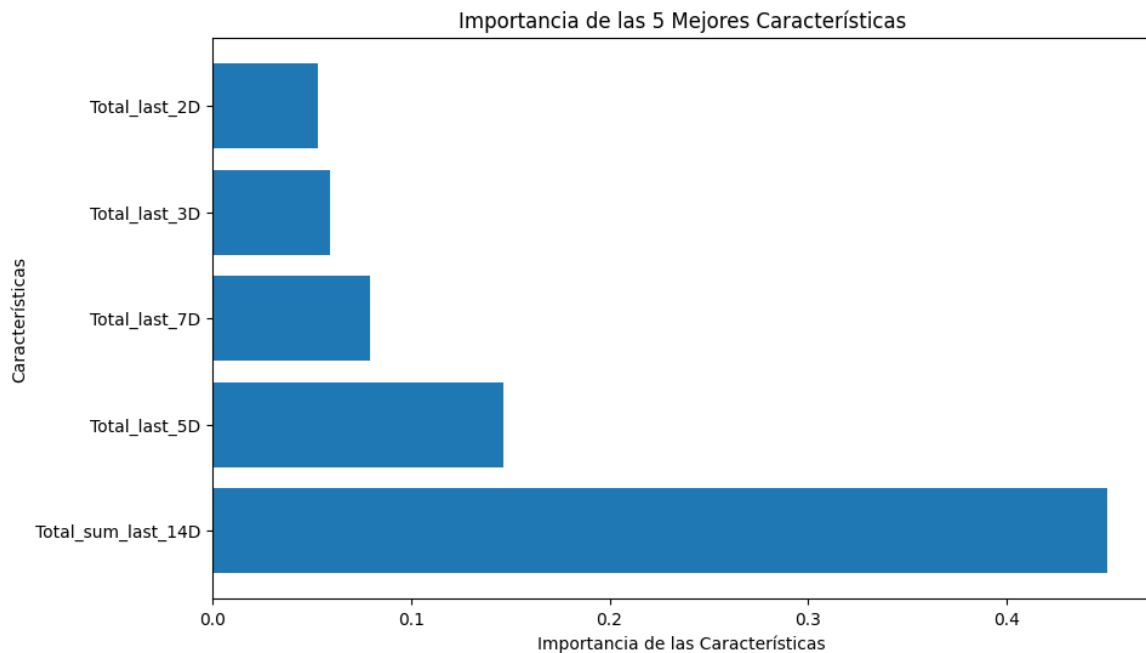


Figura 24. El gráfico de la 5 mejores características, se construyó mediante el algoritmo de Bosques Aleatorios utilizando el conjunto de datos CRUE. Se calculó la importancia de cada característica evaluando su contribución a la capacidad predictiva del modelo. En la gráfica, cada barra horizontal representa una característica, y su longitud indica la magnitud de la importancia. Las características se han etiquetado en el eje vertical. Cuanto más alta sea la barra, mayor será la importancia de la característica correspondiente. Las características destacadas son la suma total de los últimos 14 días y los últimos 5 días

La figura 25 muestra la serie temporal que abarca los conjuntos de entrenamiento, prueba y las predicciones del modelo. La línea gris representa la fase de entrenamiento del modelo, mientras que el área llamada test representa el conjunto de datos de prueba. Los marcadores verdes en forma de triángulo ilustran cómo el modelo se ajusta a la ocurrencia de eventos en comparación con los datos reales. Este gráfico permite visualizar la capacidad del modelo para capturar patrones temporales y su desempeño en el conjunto de prueba.

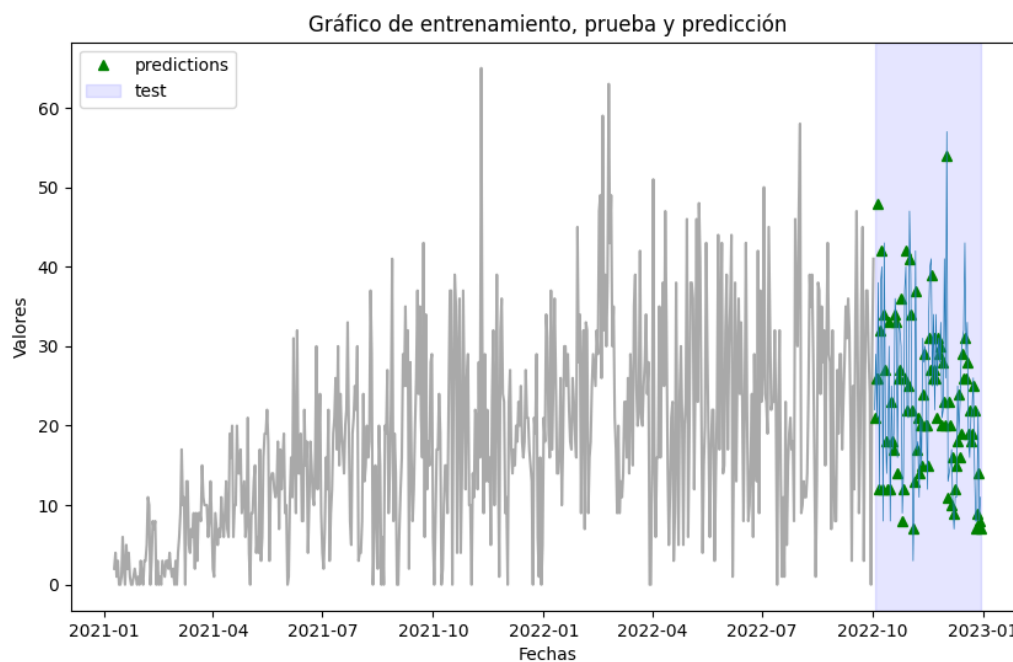


Figura 25. Serie temporal de entrenamiento, test y predicciones. conjunto de dato CRUE

Esta representación gráfica ver figura 26 se centra exclusivamente en el conjunto de prueba y las predicciones del modelo. Al enfocarse únicamente en estos dos aspectos, se facilita la comparación directa entre las observaciones reales y las anticipadas por el modelo. Este gráfico es esencial para evaluar la precisión del modelo en situaciones del mundo real, proporcionando una visión de su capacidad para prever eventos médicos de emergencia.

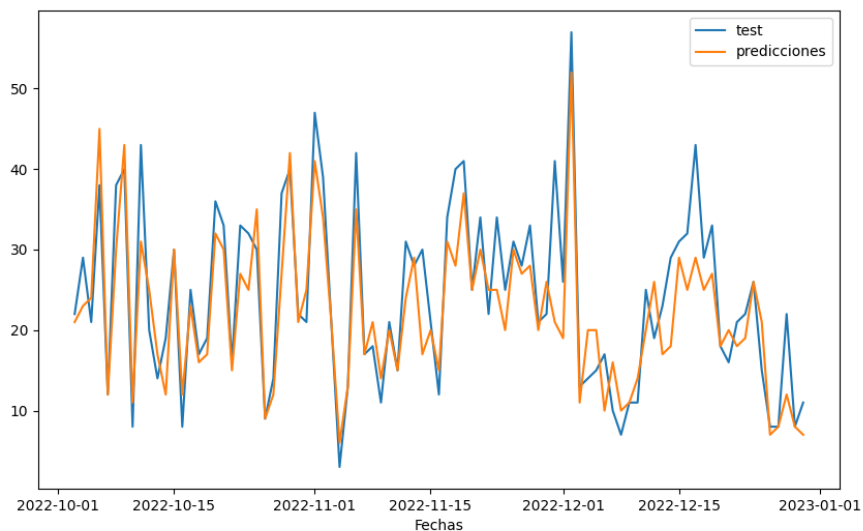


Figura 26. Serie temporal de datos acumulativos de test y predicciones, conjunto de datos CRUE

La figura 27 presenta mapas de calor que visualizan la distribución geográfica de los eventos reales de prueba y las predicciones del modelo. Cada mapa destaca la concentración de eventos en áreas específicas del área metropolitana de Bucaramanga. La comparación entre los dos mapas revela la capacidad del modelo para capturar patrones espaciales y ofrece información valiosa sobre las zonas con mayor incidencia de emergencias médicas. Este tipo de visualización es esencial para respaldar la toma de decisiones logísticas y mejorar la eficiencia en la respuesta a emergencias médicas en entornos urbanos.

Los gráficos presentados continuación son representación de los resultados a partir de conjunto de datos de la dirección de tránsito de Bucaramanga. Donde se puede observar un mejor ajuste en comparación con el modelo que hace uso de conjunto de datos de el CRUE

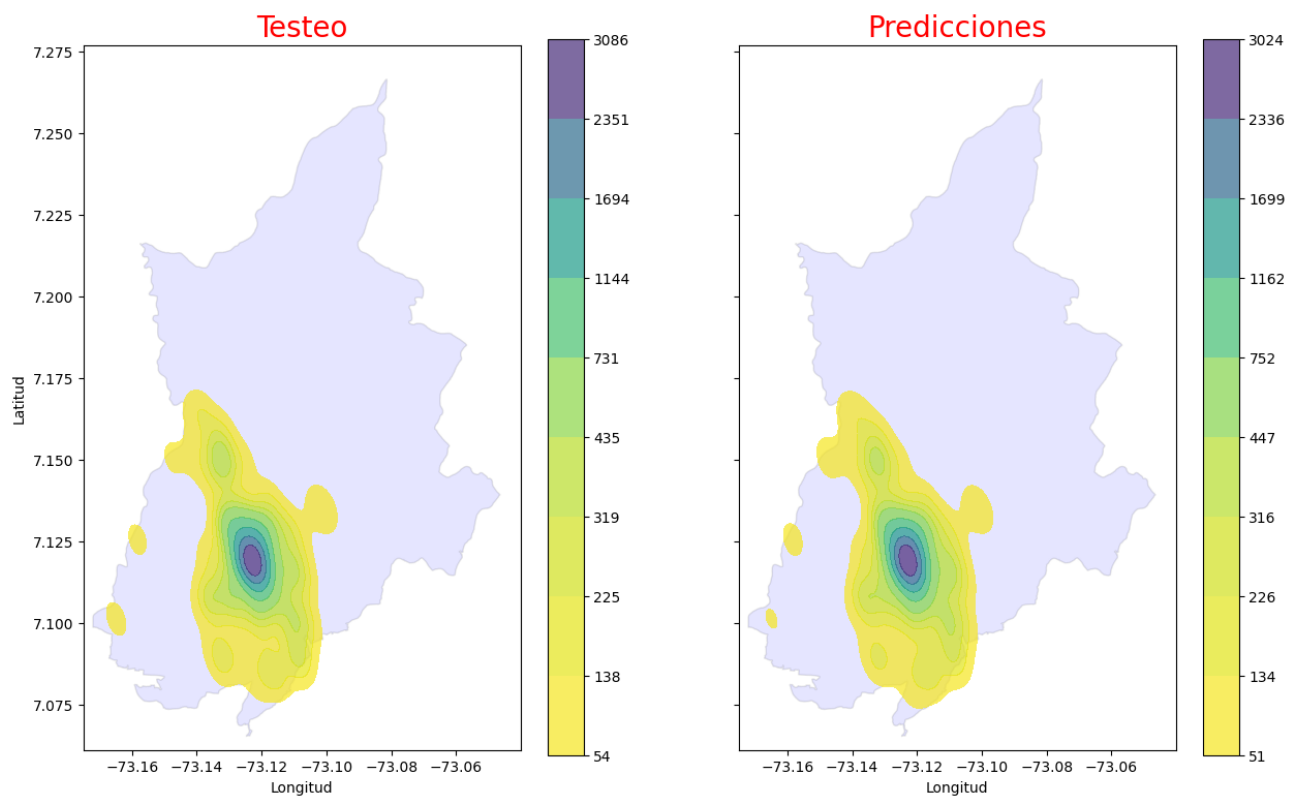


Figura 27. Mapas de calor en la zona de estudio de los datos de prueba y predicciones, del conjunto de datos CRUE

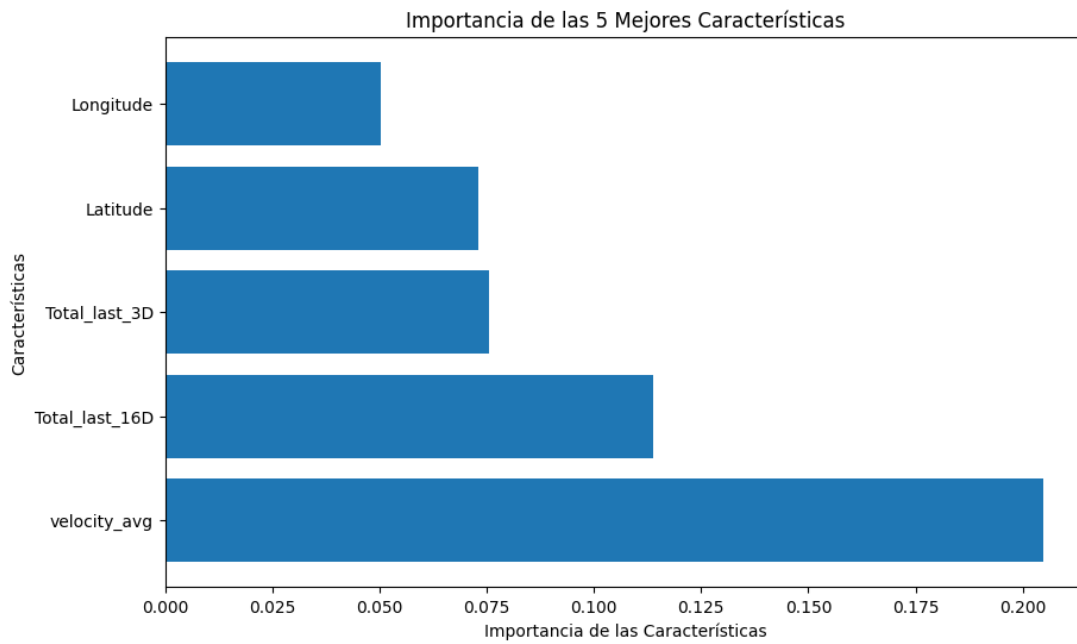


Figura 28. 5 mejores característica en el modelo de la DTB(2017-2019)

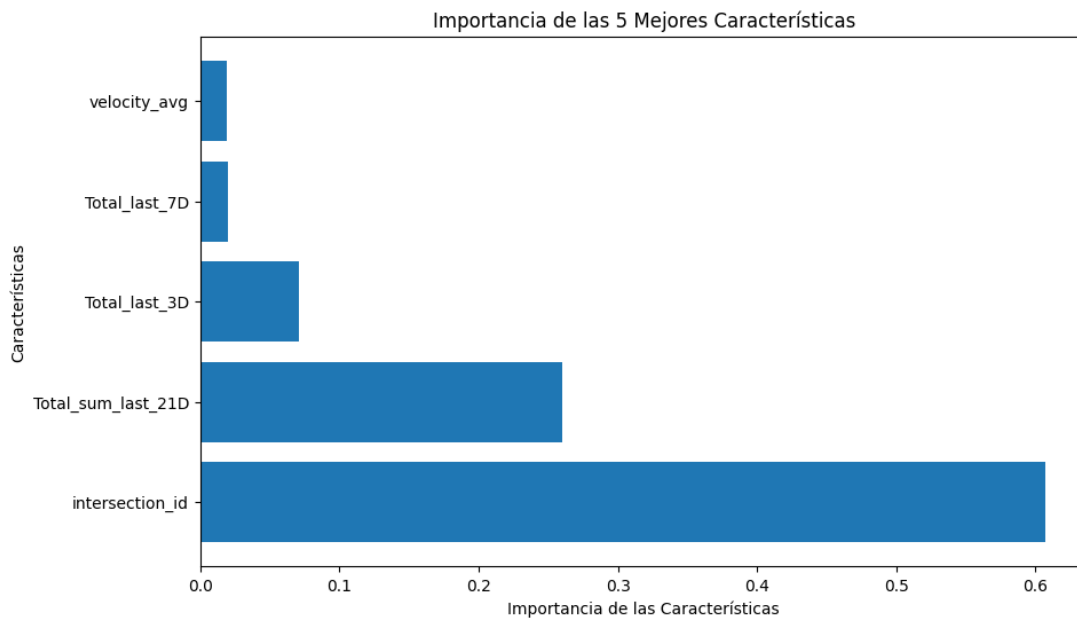


Figura 29. 5 mejores característica en el modelo de la DTB(2021-2022)

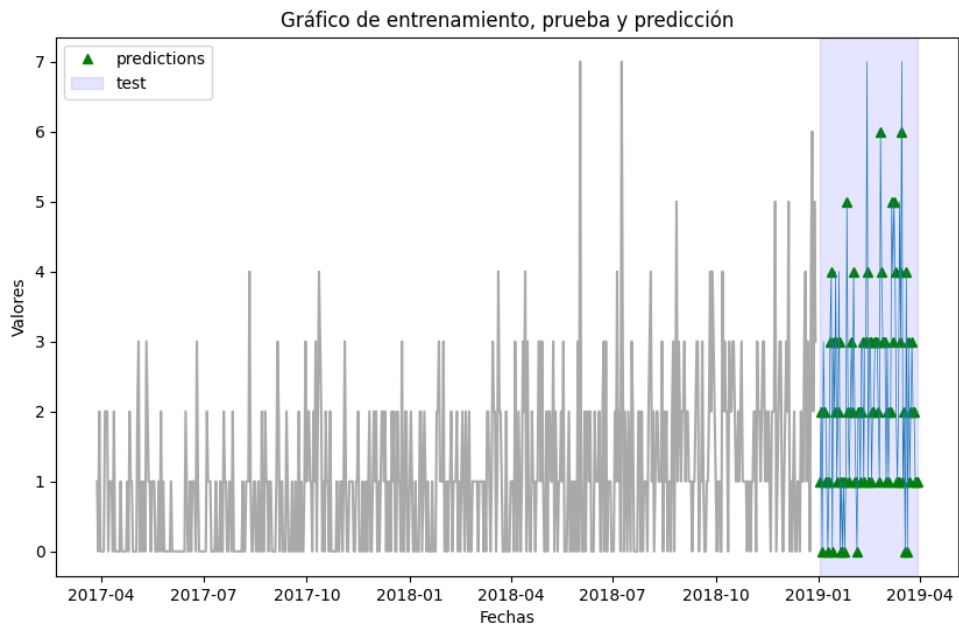


Figura 30. Serie temporal de entrenamiento, testeo y predicciones, conjunto de datos DTB(2017-2019)

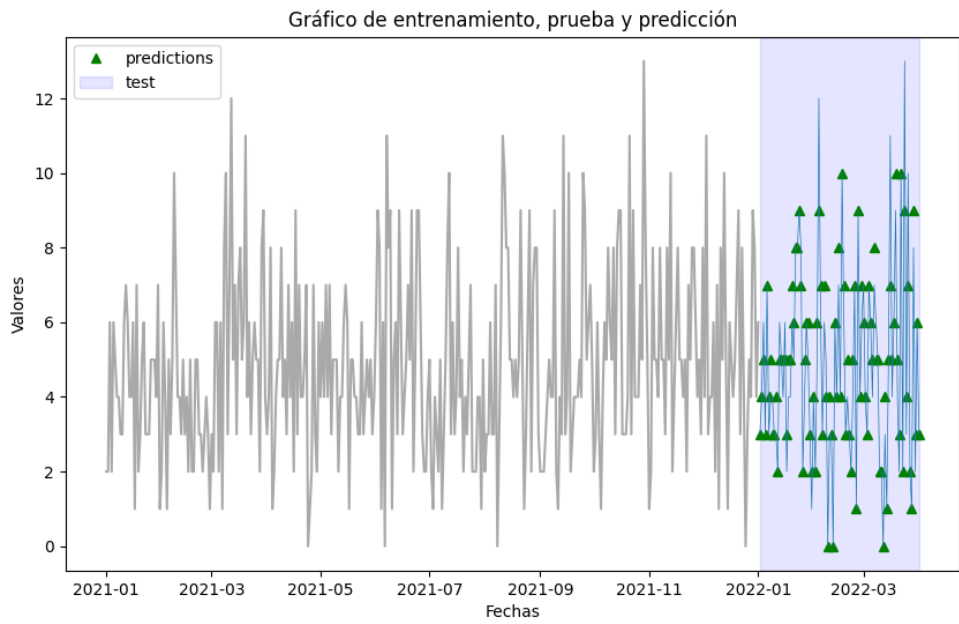


Figura 31. Serie temporal de entrenamiento, testeo y predicciones, conjunto de datos DTB(2021-2022)

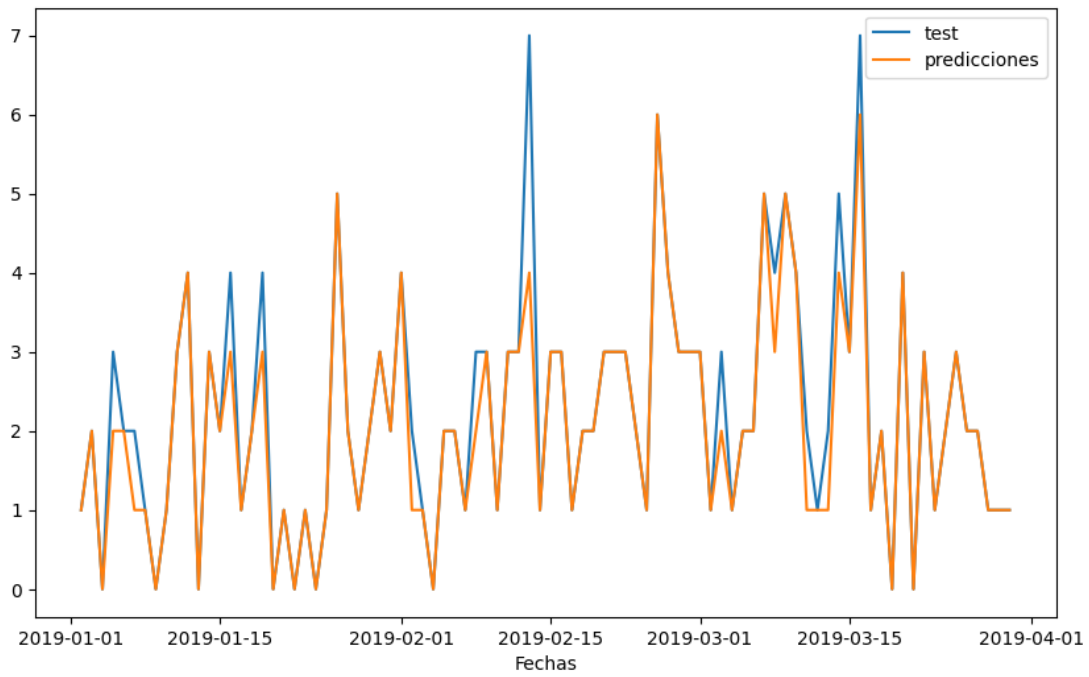


Figura 32. Serie temporal de testeo y predicciones, conjunto de datos CRUE

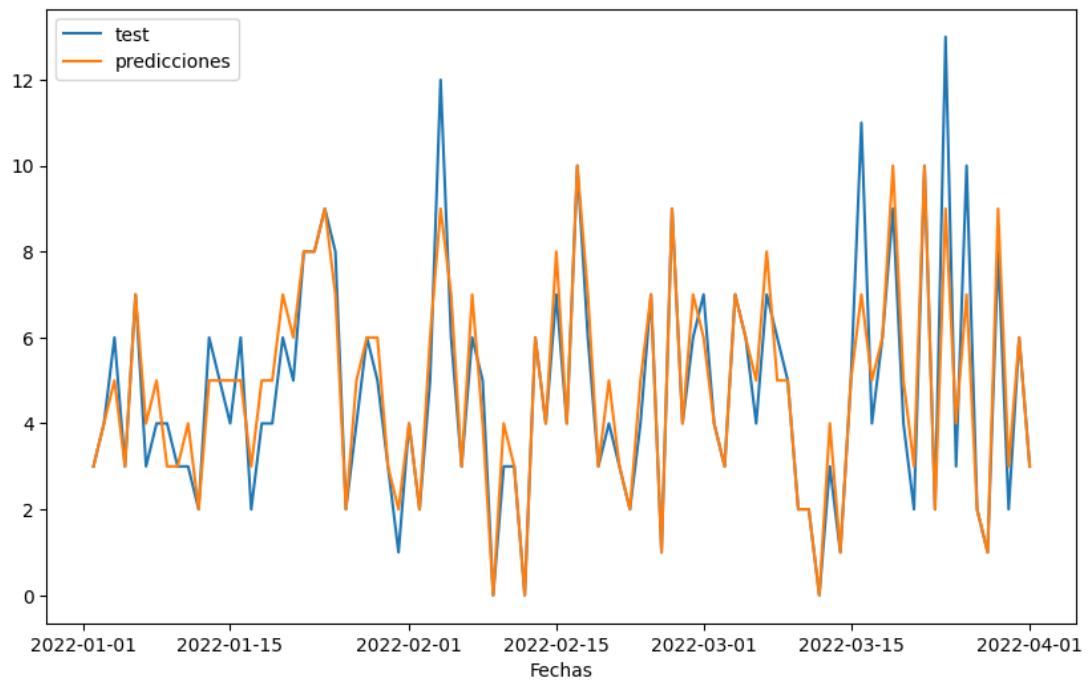


Figura 33. Serie temporal de testeo y predicciones, conjunto de datos DTB(2017-2019)

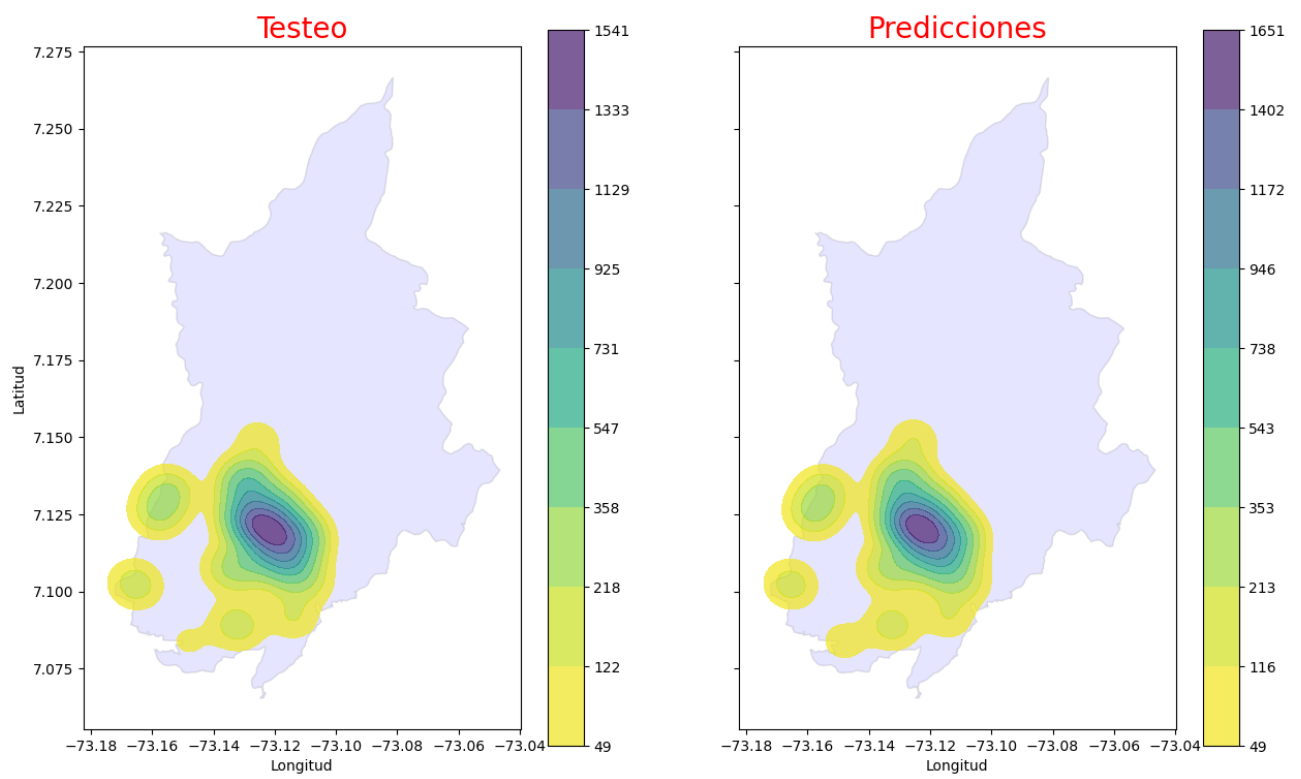


Figura 34. Mapas de calor de datos de prueba y predicciones, conjunto de datos DTB(2017-2019)

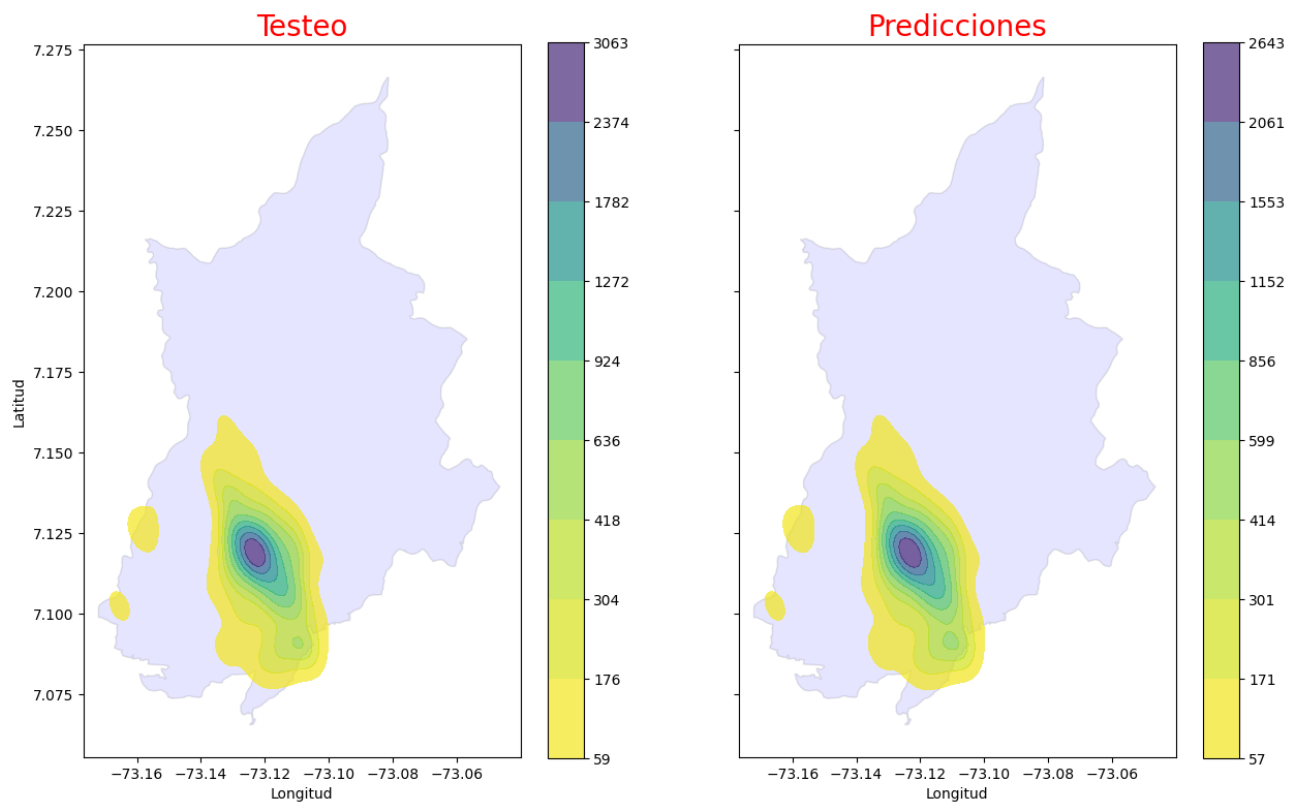


Figura 35. Mapas de calor de datos de prueba y predicciones, conjunto de datos DTB(2021-2022)

9. CONCLUSIONES Y PERSPECTIVAS

La presente investigación ha abordado de manera exhaustiva el desarrollo y evaluación de modelos predictivos para emergencias médicas y accidentes de tránsito en el área metropolitana de Bucaramanga. A lo largo de este proceso, se han explorado y aplicado diversas técnicas y metodologías para mejorar la capacidad de predicción y comprensión de estos eventos críticos.

Uno de los aspectos clave de esta investigación fue la aplicación de la metodología de ventana deslizante para la formación de vectores de características. Esta técnica demostró ser fundamental para capturar patrones temporales y espaciales relevantes, permitiendo la adaptación del modelo a la dinámica cambiante de las emergencias médicas y los accidentes de tránsito a lo largo del tiempo. La optimización de la ventana deslizante, con especial énfasis en la configuración de longitud y frecuencia de deslizamiento, no solo mejoró la capacidad predictiva de los modelos, sino que también los hizo más eficientes en términos de recursos computacionales, favoreciendo su implementación en sistemas embebidos o entornos con limitaciones de recursos.

La ingeniería de características desempeñó un papel crucial en la mejora de la representación de los datos. La incorporación de variables temporales adicionales, como el día del mes, día de la semana y semana del año, permitió una modelización más precisa de la variabilidad temporal asociada con los eventos estudiados. Además, la aplicación de representaciones cíclicas de estas variables mostró ser una estrategia efectiva para abordar la naturaleza circular de las unidades temporales, mejorando así la capacidad del modelo para capturar patrones estacionales.

La normalización de los datos fue esencial debido a las diferencias en las magnitudes de los conjuntos de datos utilizados. La evaluación de varios métodos de normalización, como Min-Max Scaling, Z-score Normalization y Robust Scaling, reveló sus impactos específicos en el rendimiento del modelo. Este proceso permitió una comparación justa y efectiva entre las diferentes variables, facilitando la interpretación de los resultados y mejorando la robustez del modelo. Los experimentos detallados, tanto con el dataset del CRUE como con el de la DTB, proporcionaron información valiosa sobre la contribución relativa de diversas configuraciones y variables al rendimiento general del modelo. Más de 40 experimentos fueron diseñados y ejecutados, lo que incluyó pruebas con distintas configuraciones de ventanas

deslizantes, evaluaciones de métodos de normalización y experimentación con variables derivadas de la ingeniería de características. Los resultados de estos experimentos fueron fundamentales para la toma de decisiones informadas sobre la configuración final del modelo.

Como aspecto relevante se puede afirmar que el ajuste del modelo seleccionado es aceptable para el conjunto de datos principal de emergencias médicas, debido a que la complejidad del problema y el área que se estudió es significativa, por lo tanto el resultado obtenido es interesante y valida la posibilidad de relacionar variables temporales y geográficas en contextos específicos como la atención de emergencias médicas.

BIBLIOGRAFÍA

- Adams, Anthony y Peter Vamplew. “Encoding and Decoding Cyclic Data”. En: *The South Pacific Journal of Natural Science* 16 (1 de ene. de 1998) (vid. pág. 66).
- Agencia Nacional de Seguridad Vial. *Histórico víctimas / ANSV*. Histórico víctimas | ANSV. URL: <https://ansv.gov.co/es/observatorio/estad%C3%ADsticas/historico-victimas> (visitado 22-08-2023) (vid. pág. 14).
- Ayodele, Taiwo. “Types of Machine Learning Algorithms”. En: *New Advances in Machine Learning*. 1 de feb. de 2010. DOI: 10.5772/9385 (vid. págs. 24, 25).
- Boeing, Geoff. “OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks”. En: *Computers, Environment and Urban Systems* 65 (1 de sep. de 2017), págs. 126-139. DOI: 10.1016/j.compenvurbsys.2017.05.004 (vid. págs. 46, 64).
- Breiman, Leo. “Random Forests”. En: *Machine Learning* 45.1 (1 de oct. de 2001), págs. 5-32. DOI: 10.1023/A:1010933404324 (vid. págs. 30, 77, 78).
- Brockwell, Peter J. y Richard A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. New York, NY: Springer, 1991. DOI: 10.1007/978-1-4419-0320-4 (vid. pág. 31).
- Castrejon, Rey Francisco Ayala y Christian Bucio Pacheco. “Modelo ARIMA aplicado al tipo de cambio peso-dólar en el periodo 2016-2017 mediante ventanas temporales deslizantes”. En: *Revista Mexicana de Economía y Finanzas Nueva Época REMEF* 15.3 (30 de jun. de 2020). Number: 3, págs. 331-354. DOI: 10.21919/remef.v15i3.466 (vid. pág. 72).

Chasco, Coro. “Geodaspace: A Resource for Teaching Spatial Regression Models”. En: *Rect@ 4* (2013). Num Pages: 26 Place: Valencia, Spain Publisher: Ramón Sala Garrido, págs. 119-144 (vid. pág. 28).

Chicco, Davide, Matthijs J. Warrens y Giuseppe Jurman. “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation”. En: *PeerJ. Computer Science* 7 (2021), e623. DOI: 10.7717/peerj-cs.623 (vid. pág. 80).

Congreso de Colombia. *Ley 2161 de 2021*. 26 de nov. de 2021 (vid. pág. 61).

Departamento Administrativo Nacional de Estadística (DANE). *Censo Nacional de Población y Vivienda 2018*, DANE. 10 de mayo de 2019 (vid. págs. 41, 47).

— *Marco Geoestadístico Nacional Nivel Municipio 2022*. Ver. MGN2022-Nivel Municipio. 7 de dic. de 2022 (vid. págs. 44, 55).

— *Marco Geoestadístico Nacional Sección Urbana 2022*. Ver. MGN2022-Nivel Sección urbano. 7 de dic. de 2022 (vid. pág. 45).

— *Marco Geoestadístico Nacional Sección Urbana integrado al CNPV2018*. Ver. Versión MGN_ANM2018-Sección Urbana. 27 de dic. de 2021 (vid. pág. 47).

— *Uso del marco geoestadístico nacional*. Sep. de 2020 (vid. págs. 44, 45).

Devangela. *googlemaps*. <https://pypi.org/project/googlemaps/>. Ver. 4.10.0. 26 de ene. de 2023 (vid. pág. 55).

Dirección de Tránsito de Bucaramanga. *Nuestra Entidad*. DTB - Dirección de Tránsito de Bucaramanga. URL: <https://transitobucaramanga.gov.co/dtb/quienes-somos/nuestra-entidad/> (visitado 31-08-2023) (vid. pág. 42).

Díaz-Hierro, J. et al. “Evaluación de modelos de series temporales para la previsión de la demanda de emergencias sanitarias”. En: *Emergencias (St. Vicenç dels Horts)* (2012), págs. 181-188 (vid. pág. 34).

Equipo desarrollador QGIS. *QGIS*. Ver. 3.32.2. 23 de jun. de 2023 (vid. pág. 15).

Esri. *ArcGIS*. Ver. 10.8.2. 9 de dic. de 2021 (vid. pág. 15).

Florax, Raymond J.G.M. y Peter Nijkamp. “Misspecification in Linear Spatial Regression Models”. En: *Tinbergen Institute Discussion Papers 2003-081/3* (3 de ago. de 2003) (vid. pág. 29).

Galli, Soledad. “Feature-engine: A Python package for feature engineering for machine learning”. En: *Journal of Open Source Software* 6.65 (2021), pág. 3642. DOI: 10.21105/joss.03642 (vid. pág. 67).

García, Andrés Camilo Florez y José Nelson Pérez Castillo. “Técnicas para la predicción espacial de zonas susceptibles a deslizamientos”. En: *Avances Investigación en Ingeniería* 16.1 (21 de ago. de 2019). Number: 1, págs. 20-48. DOI: 10.18041/1794-4953/avances.1.5188 (vid. pág. 32).

Google Maps. *Geocoding API*. <https://developers.google.com/maps/documentation/geocoding/overview?hl=es> 419. 2023 (vid. pág. 55).

— *Routes API*. <https://developers.google.com/maps/documentation/routes?hl=es> 419. 2023 (vid. págs. 53, 69).

- Grekousis, George. “Artificial neural networks and deep learning in urban geography: A systematic review and meta-analysis”. En: *Computers, Environment and Urban Systems* 74 (2019), págs. 244-256. DOI: <https://doi.org/10.1016/j.compenvurbsys.2018.10.008> (vid. pág. 30).
- Guesgen, Hans W. et al. “Buffering Fuzzy Maps in GIS”. En: *Spatial Cognition & Computation* 3.2 (1 de sep. de 2003). Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13875868.2003.9683762> págs. 207-222. DOI: 10.1080/13875868.2003.9683762 (vid. pág. 65).
- Gutiérrez Corea, Federico Vladimir. “Predicción espacio-temporal de la irradiancia solar global a corto plazo en España mediante geoestadística y redes neuronales artificiales”. <http://purl.org/dc/dmity> Universidad Politécnica de Madrid, 2014 (vid. pág. 33).
- Hagberg, Aric, Pieter Swart y Daniel S Chult. *Exploring network structure, dynamics, and function using NetworkX*. Inf. téc. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008 (vid. pág. 64).
- Huang S. Cai N., Penzuti Pacheco P. Narrandes S. Wang Y. y Xu W. “Applications of Support Vector Machine (SVM) Learning in Cancer Genomics”. En: *Cancer Genomics - Proteomics* 15 (ene. de 2018) (vid. pág. 30).
- Hunter, J. D. “Matplotlib: A 2D graphics environment”. En: *Computing in Science & Engineering* 9.3 (2007), págs. 90-95. DOI: 10.1109/MCSE.2007.55 (vid. págs. 46, 59, 64).
- Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM. *Datos de precipitación de pluviómetro convencional en Santander, Bucaramanga*. 16 de dic. de 2023 (vid. pág. 47).
- Jordahl, Kelsey et al. *geopandas/geopandas: v0.8.1*. Ver. v0.8.1. Jul. de 2020. DOI: 10.5281/zenodo.3946761 (vid. págs. 46, 56).

Lamprecht, Christian. *Meteostat*. Ver. 1.6.7. Germany, 28 de sep. de 2023 (vid. pág. 49).

Layton, Robert y Paul A. Watters. *Automating Open Source Intelligence: Algorithms for OSINT*. San Diego, UNITED STATES: Elsevier Science & Technology Books, 2015 (vid. pág. 28).

Le X-H. Ho HV., Lee G. y Jung S. “Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting.” En: *Water* 11 (5 de jul. de 2019) (vid. pág. 32).

Li, Jin et al. “Application of machine learning methods to spatial interpolation of environmental variables”. En: *Environmental Modelling & Software* 26.12 (dic. de 2011), págs. 1647-1659. DOI: 10.1016/j.envsoft.2011.07.004 (vid. pág. 78).

Martijn Vermaat y otros. *Calmap: Calendar heatmaps from Pandas time series data*. Ver. 0.0.11. 31 de jul. de 2023 (vid. pág. 51).

“Mean Squared Error”. En: *The Concise Encyclopedia of Statistics*. New York, NY: Springer, 2008, págs. 337-339. DOI: 10.1007/978-0-387-32833-1_251 (vid. pág. 81).

Ministerio de salud y proteccion social. *Resolucion Número 926 de 2017*. 30 de mar. de 2017 (vid. págs. 14, 17, 21, 37).

— *Sistema de emergencias médicas SEM*. Sistema de emergencias medicas-SEM. 29 de dic. de 2019. URL: <https://www.minsalud.gov.co/salud/PServicios/Paginas/Sistema-de-emergencias-medicas-SEM.aspx> (visitado 23-08-2023) (vid. págs. 19, 37).

Ministerio de Salud y Protección Social. *Gestión integral de riesgo en salud*. Gestión integral de riesgo en salud. 19 de dic. de 2019. URL: <https://www.minsalud.gov.co/proteccionsocial/Paginas/gestion-integral-de-riesgo-en-salud.aspx> (visitado 29-08-2023) (vid. pág. 36).

Ministerio de Salud y Protección Social. *Registro Especial de Prestadores de Servicios de Salud (REPS)*. 10 de dic. de 2023 (vid. pág. 42).

— *Resolución Número 1220 de 2010*. 8 de abr. de 2010 (vid. pág. 42).

Ministerio de Salud y Protección Social, Instituto Nacional de Salud y Observatorio Nacional de Salud. *Aspectos relacionados con la frecuencia de uso de los servicios de salud, mortalidad y discapacidad en Colombia*. 2011 (vid. pág. 22).

Nelson, Brian K. “Time Series Analysis Using Autoregressive Integrated Moving Average (ARIMA) Models.” En: *Academic Emergency Medicine* 5 (28 de jun. de 2008) (vid. pág. 31).

OpenStreetMap contributors. *Planet dump retrieved from <https://planet.osm.org>*. <https://www.openstreetmap.org>. 2017 (vid. págs. 46, 56).

Organización Mundial de la Salud. *Servicios sanitarios de calidad*. Servicios sanitarios de calidad. 11 de ago. de 2020. URL: <https://www.who.int/es/news-room/fact-sheets/detail/quality-health-services> (visitado 22-08-2023) (vid. pág. 15).

P., Bostan. “Basic Kriging Methods in Geostatistics”. En: *Yuzuncu Yil University Journal of Agricultural Sciences* 27 (31 de mar. de 2017) (vid. pág. 29).

Polasek, Wolfgang. *The Hodrick-Prescott (HP) filter as a Bayesian regression model*. Inf. téc. Reihe Ökonomie/Economics Series, 2011 (vid. pág. 61).

Presidente de la república. *Decreto 998 de 2022*. 13 de jun. de 2022 (vid. pág. 61).

Pérez Machado, Reinaldo. *Proceso de geocodificación urbana: ejemplos de su aplicación en la ciudad de Barcelona*. Vol. XIII. Journal Abbreviation: [RCG] Revista Catalana de Geografia. Revista digital de geografia, cartografia i ciències de la Terra Publication Title: [RCG] Re-

- vista Catalana de Geografia. Revista digital de geografia, cartografia i ciències de la Terra. 29 de mayo de 2008. DOI: 10.13140/2.1.1858.5280 (vid. pág. 54).
- Quinlan, J. R. “Induction of decision trees”. En: *Machine Learning* 1.1 (1 de mar. de 1986), págs. 81-106. DOI: 10.1007/BF00116251 (vid. pág. 76).
- Robert B. Cleveland William S. Cleveland, Jean E. McRae e Irma Terpenning. “STL: A Seasonal-Trend Decomposition Procedure Based on Loess”. En: *Journal of Official Statistics* 6 (mar. de 1990) (vid. pág. 31).
- Seabold, Skipper y Josef Perktold. “statsmodels: Econometric and statistical modeling with python”. En: *9th Python in Science Conference*. 2010 (vid. pág. 59).
- Semana. “Ambulancias sin límite: la ‘guerra del centavo’ continúa y está dejando víctimas mortales”. En: *Semana.com Últimas Noticias de Colombia y el Mundo* (18 de jun. de 2022). Section: Enfoque (vid. pág. 42).
- Shen J., Valagolam D. y McCalla S. “Prophet forecasting model: a machine learning approach to predict the concentration of air pollutants (PM2.5, PM10, O3, NO2, SO2, CO) in Seoul, South Korea”. En: *PeerJ* (15 de sep. de 2020) (vid. pág. 32).
- Swistakm. *gmaps*. <https://pypi.org/project/gmaps/>. Ver. latest (vid. pág. 55).
- Tableau Software. *Tableau Prep Builde*. Ver. 2022.2 (vid. pág. 54).
- team, The pandas development. *pandas-dev/pandas: Pandas*. Ver. latest. Feb. de 2020. DOI: 10.5281/zenodo.3509134 (vid. págs. 49, 55).
- Universidad de Antioquia y Ministerio de Salud y Protección Social. *Guías Básicas de Atención Médica Prehospitalaria*. 2012 (vid. págs. 19, 20).

visualization, python. *Folium*. Ver. 0.11.0. 28 de dic. de 2020 (vid. pág. 56).

Zambrano, Rafael. “Un enfoque espaciotemporal para la predicción de delitos en la ciudad de buenos aires”. En: *Revista de investigación en modelos matemáticos aplicados a la gestión y la economía* 2.7 (13 de nov. de 2021) (vid. págs. 34, 64).

Área Metropolitana de Bucaramanga. *Bucaramanga*. AMB. 10 de abr. de 2012. URL: <https://www.amb.gov.co/bucaramanga/> (visitado 31-08-2023) (vid. pág. 41).

ANEXOS

Anexo A. Función para realizar peticiones utilizando la API de Google RUTES, implementada mediante python

Código para el manejo de respuestas

```
import requests
import json
from models.ResponseModel import Response

class RouteAPI:
    def __init__(self, api_key):
        self.api_key = api_key
        self.base_url = 'https://routes.googleapis.com/directions/v2:computeRoutes'

    def make_api_request(self, origin_lat, origin_lng, destination_lat,
        destination_lng):
        headers = {
            'Content-Type': 'application/json',
            'X-Goog-API-Key': self.api_key,
            'X-Goog-FieldMask': (
                'routes.duration,routes.distanceMeters,routes.legs.polyline,'
                'routes.legs.travelAdvisory,routes.legs.steps.polyline,'
                'routes.legs.steps.distanceMeters,routes.legs.steps.staticDuration,'
                'routes.legs.startLocation,routes.legs.endLocation'
            )
        }
    }
```

```

data = {
    "origin": {
        "location": {
            "latLng": {
                "latitude": origin_lat,
                "longitude": origin_lng
            }
        }
    },
    "destination": {
        "location": {
            "latLng": {
                "latitude": destination_lat,
                "longitude": destination_lng
            }
        }
    },
    "travelMode": "DRIVE",
    "units": "METRIC",
    "languageCode": "es-CO",
    "extraComputations": ["TRAFFIC_ON_POLYLINE"],
    "routingPreference": "TRAFFIC_AWARE",
    "polylineEncoding": "GEO_JSON_LINestring"
}

try:
    response = requests.post(self.base_url, headers=headers,
                              data=json.dumps(data))
    if response.status_code == 200:

```

```
    response_json = response.json()
    return Response(success=True, data=response_json)
else:
    error_message = f"Error in API request. Status code:
                    {response.status_code}"
    return Response(success=False, error_message=error_message)
except requests.exceptions.RequestException as e:
    error_message = f"Error in API request. Status code: {e}"
    return Response(success=False, error_message=error_message)
```

Anexo B. Respuesta en formato JSON de la simulación de rutas específica para la extracción de variables relevante.

Estructura de respuesta.

```
{
  "routes": [
    {
      "legs": [
        {
          "polyline": {
            "geoJsonLinestring": {
              "type": "LineString",
              "coordinates": [
                [-73.1157965, 7.1273200999999995],
                [-73.11621880459006, 7.127074428116561],
                [-73.1159521, 7.1265385000000006],
                [-73.1156676, 7.1260196999999999],
                [-73.1150902, 7.1263473],
                [-73.1141659, 7.1268397],
                [-73.11336, 7.1271866999999999],
                [-73.11300456542745, 7.127314468639382],
                [-73.11295352541131, 7.1273331230979045],
                [-73.1142821, 7.1297736],
                [-73.1144242, 7.1299449],
                [-73.11450529999999, 7.1300338],
                [-73.11461369999999, 7.1300674],
                [-73.11472169999999, 7.1306522999999995],
                [-73.11509439999999, 7.1323158999999998],
                [-73.1152263, 7.1331023],
```

```

        [-73.1153915, 7.133956599999999],
        [-73.11550940000001, 7.134371399999999],
        [-73.115337, 7.1343646]
    ]
}
},
"startLocation": {
    "latLng": {
        "latitude": 7.1273200999999995,
        "longitude": -73.1157965
    }
},
"endLocation": {
    "latLng": { "latitude": 7.1343646, "longitude": -73.115337 }
},
"steps": [
    {
        "distanceMeters": 54,
        "staticDuration": "20s",
        "polyline": {
            "geoJsonLinestring": {
                "type": "LineString",
                "coordinates": [
                    [-73.1157965, 7.1273200999999995],
                    [-73.1162204, 7.1270735]
                ]
            }
        }
    }
},

```

```

{
  "distanceMeters": 132,
  "staticDuration": "38s",
  "polyline": {
    "geoJsonLinestring": {
      "type": "LineString",
      "coordinates": [
        [-73.1162204, 7.1270735],
        [-73.1159521, 7.1265385000000006],
        [-73.1156676, 7.1260197]
      ]
    }
  }
},
{
  "distanceMeters": 328,
  "staticDuration": "83s",
  "polyline": {
    "geoJsonLinestring": {
      "coordinates": [
        [-73.1156676, 7.1260197],
        [-73.1150902, 7.1263473],
        [-73.1141659, 7.1268397],
        [-73.11336, 7.127186699999999],
        [-73.113004199999999, 7.1273146]
      ],
      "type": "LineString"
    }
  }
}

```

```

},
{
  "distanceMeters": 339,
  "staticDuration": "77s",
  "polyline": {
    "geoJsonLinestring": {
      "type": "LineString",
      "coordinates": [
        [-73.11300419999999, 7.1273146],
        [-73.1129514, 7.1273339],
        [-73.1142821, 7.1297736],
        [-73.1144242, 7.1299449]
      ]
    }
  }
},
{
  "distanceMeters": 515,
  "staticDuration": "91s",
  "polyline": {
    "geoJsonLinestring": {
      "type": "LineString",
      "coordinates": [
        [-73.1144242, 7.1299449],
        [-73.11450529999999, 7.1300338],
        [-73.11461369999999, 7.1300674],
        [-73.11472169999999, 7.1306522999999995],
        [-73.11509439999999, 7.1323159],
        [-73.1152263, 7.1331023],

```

```

        [-73.1153915, 7.133956599999999],
        [-73.11550940000001, 7.1343714]
    ]
}
}
},
{
    "distanceMeters": 19,
    "staticDuration": "6s",
    "polyline": {
        "geoJsonLinestring": {
            "type": "LineString",
            "coordinates": [
                [-73.11550940000001, 7.1343714],
                [-73.115337, 7.1343646]
            ]
        }
    }
}
},
],
"travelAdvisory": {
    "speedReadingIntervals": [
        {
            "startPolylinePointIndex": 0,
            "endPolylinePointIndex": 1,
            "speed": "SLOW"
        },
        {
            "startPolylinePointIndex": 1,

```

```
        "endPolylinePointIndex": 7,  
        "speed": "NORMAL"  
    },  
    {  
        "startPolylinePointIndex": 7,  
        "endPolylinePointIndex": 8,  
        "speed": "SLOW"  
    },  
    {  
        "startPolylinePointIndex": 8,  
        "endPolylinePointIndex": 18,  
        "speed": "NORMAL"  
    }  
] ]  
}  
],  
"distanceMeters": 1387,  
"duration": "215s"  
}  
]  
}
```
