

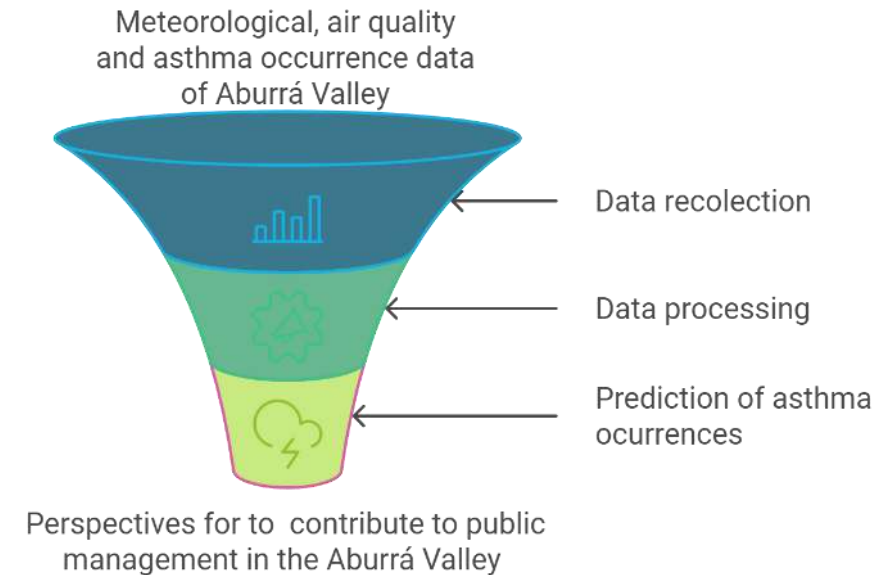
Modelo computacional de Inteligencia Artificial para la Predicción de Exacerbaciones de Asma, en poblaciones expuestas a condiciones de calidad del aire desfavorables

Yesid O. Ramírez C; Juan D. Lemos D; Mauricio A. Correa O.



Introducción

El proyecto busca a través de la implementación de herramientas de **DeepLearning**, procesar datos meteorológicos, de calidad del aire y de ocurrencia de asma para **predecir comportamientos de ocurrencia de asma** de acuerdo a datos meteorológicos y de calidad del aire en el Valle de Aburrá, y **contribuir** a la **gestión pública** en el Valle de Aburrá con un modelo de IA.



Contexto del problema

Topografía irregular y geografía del Valle de Aburrá



Meteorología

Nubes de baja altura
Ventilación escasa
Clima cambiante,
propio de una región tropical

Topografía

Rodeado de montañas
Valle estrecho,
densamente poblado

Emisión de contaminantes

Producto de actividades humanas:

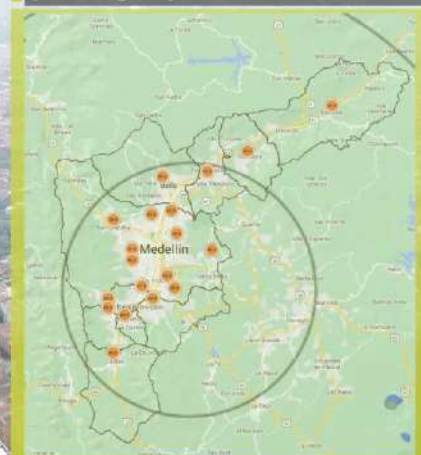
Fuentes fijas : producción industrial

Fuentes móviles : transporte automotor

Planteamiento del problema

La Necesidad de herramientas **predictivas** para mitigar el impacto de la contaminación sobre la salud respiratoria en el Valle de Aburrá

El 5 de marzo de 2024, las 19 estaciones estaban en color naranja y reportaban que el aire era dañino para grupos sensibles





■ Impactos de la mala calidad del aire

- El **12,3% (2105)** del total de las muertes que se presentan en el área, son atribuidas a la contaminación del aire urbano, **generó costos** estimados en **\$2,8 billones de pesos**, equivalentes al 5% del PIB del área.[1]
- **Deterioro de la calidad de vida**, costos en salud, pérdida de productividad. [2]
- Aumento de **gastos en control de limpieza**, se desvían recursos del desarrollo social [2]
- **Disminuye la satisfacción** con el entorno, menos oportunidades socioculturales, recreativas y deportivas. [2]
- **Contaminación** del aire, agua y suelos, pérdida de biodiversidad, presión sobre ecosistemas y recursos hídricos [3]

Contexto y justificación del problema

¿POR QUÉ ASMA?

- Su contención tiene un **mayor impacto socioeconómico**, debido a la carga pediátrica y de urgencias que conlleva. (Asma leve : \$502.752 COP - Asma moderada : \$769.454 COP - Asma grave: \$5.924.398 COP, datos tarifarios SOAT, ISS [4])
- **1 de cada 8 colombianos padece asma** [5] - En Antioquia 16% la prevalencia de asma a 2019 [6] [9] - Según la OMS, se estima que 262 millones de personas tuvieron asma en 2019 y que esta enfermedad causó 455000 muertes.
- Grupo etario más amplio, más **sensible a las variables climáticas** y de calidad del aire para una mejor modelización.
- Afección crónica, puede **coexistir** con otras **enfermedades**.
- Se genera un **modelo más específico**, preciso y **robusto** en relación con el asma
- **Ausencias** laborales y escolares debido a incapacidad.

Objetivos

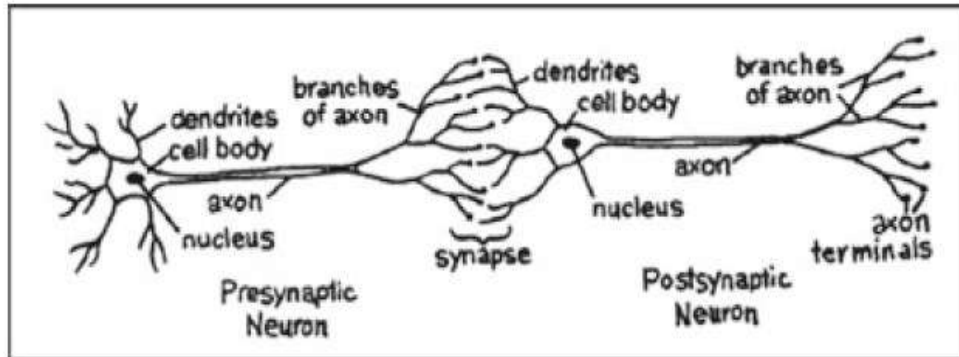
General

Desarrollar un modelo de **predicción de** ocurrencias de asma a partir de la información de calidad el aire, meteorología para generar alertas que contribuyan en la **gestión sanitaria** del Valle de Aburrá

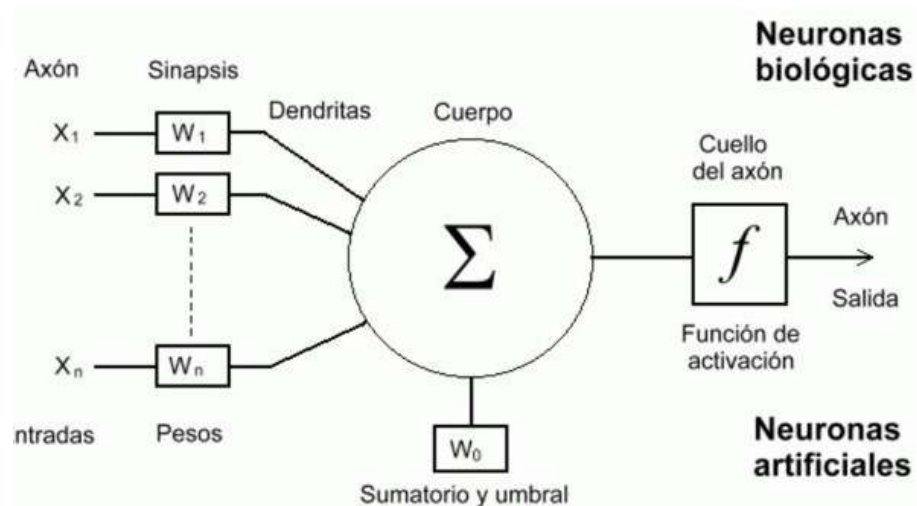
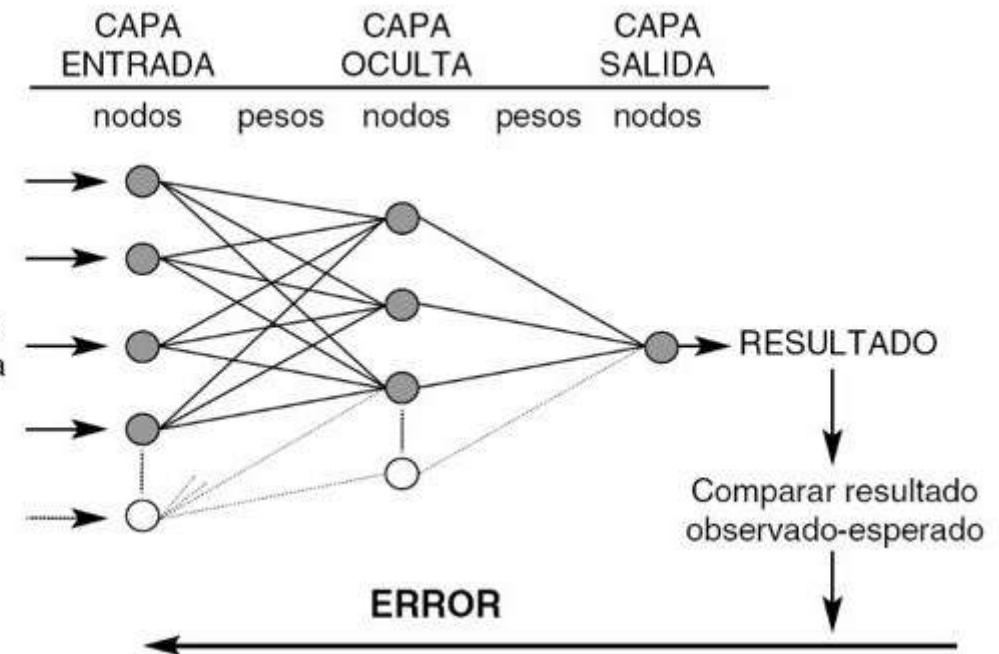
Específicos

- Analizar enfoques basados en **inteligencia artificial (IA)** aplicados a la predicción de eventos de salud respiratoria asociados a condiciones climáticas y ambientales
- Seleccionar y estructurar **datos** relevantes sobre **asma , calidad del aire y variables meteorológicas** , integrando fuentes sanitarias y de monitoreo ambiental
- Desarrollar y ajustar un **modelo LSTM** que permita predecir exacerbaciones asmáticas, considerando la naturaleza **multivariantes y temporal** de los datos.

Marco teórico y estado del arte - Redes neuronales

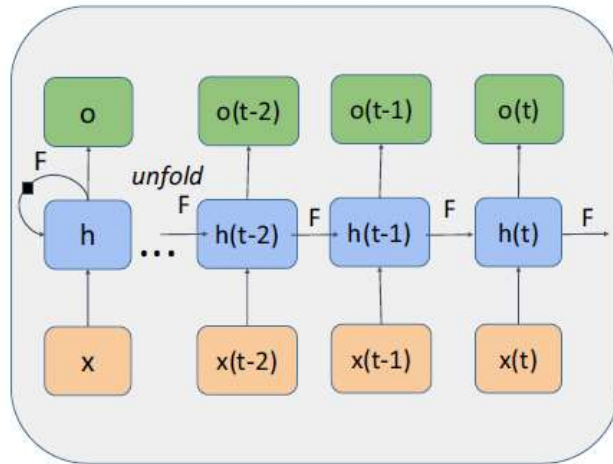


Fundamentos de las redes neuronales feedforward. [7]

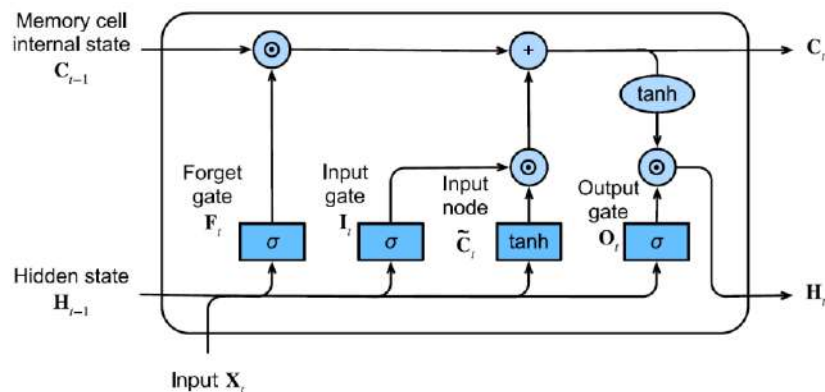
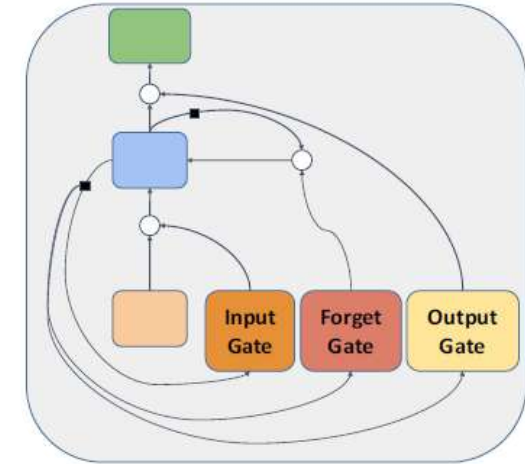


Marco teórico y estado del arte - Redes neuronales Recurrentes – Long Short Term Memory (LSTM)

Despliegue de las redes neuronales recurrentes. [8] [9]



Extensiones de las redes recurrentes con unidades de compuerta (Gate). El cuadrado negro representa un retardo de un paso de tiempo discreto [8]



$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f)$$

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i)$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o)$$

$$\tilde{C}_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c)$$

Marco teórico y estado del arte – Modelos mas usados

Redes neuronales

RNN: Red neuronal recurrente

LSTM: Long Short-Term Memory , tipo de RNN

DNN: Red neuronal profunda con retroalimentación [10]

ANN: Red neuronal artificial

Modelos híbridos de ML y DL

- RNN – LSTM [11]

- VAR – DNN [12]

- DT – ANN [13]

Metodología

Revisión Bibliográfica

Realización de búsquedas exhaustivas en bases de datos académicas



Procesamiento de Datos

Limpieza y normalización de datos relacionados con el asma, datos ambientales y contaminación



Selección de Metodologías

Evaluación de modelos predictivos como DNNs, ARIMA, GRU y LSTM



Implementación y Validación

Optimización y validación de modelos, utilizando Python y bibliotecas de IA



Resultados

Variables de estudio

Para propósitos del proyecto las variables de estudio fueron:

Variable objetivo (target):

Casos de asma en diferentes grados, crisis , hospitalizaciones y exacerbaciones en diferentes grupos etarios.

Variables predictoras:

- Concentraciones de PM2.5 y PM10
- Concentraciones de O3, CO, NO, NO2 y SO2
- Velocidad el viento
- Temperatura
- Humedad relativa
- Presión atmosférica
- Precipitación

Resultados

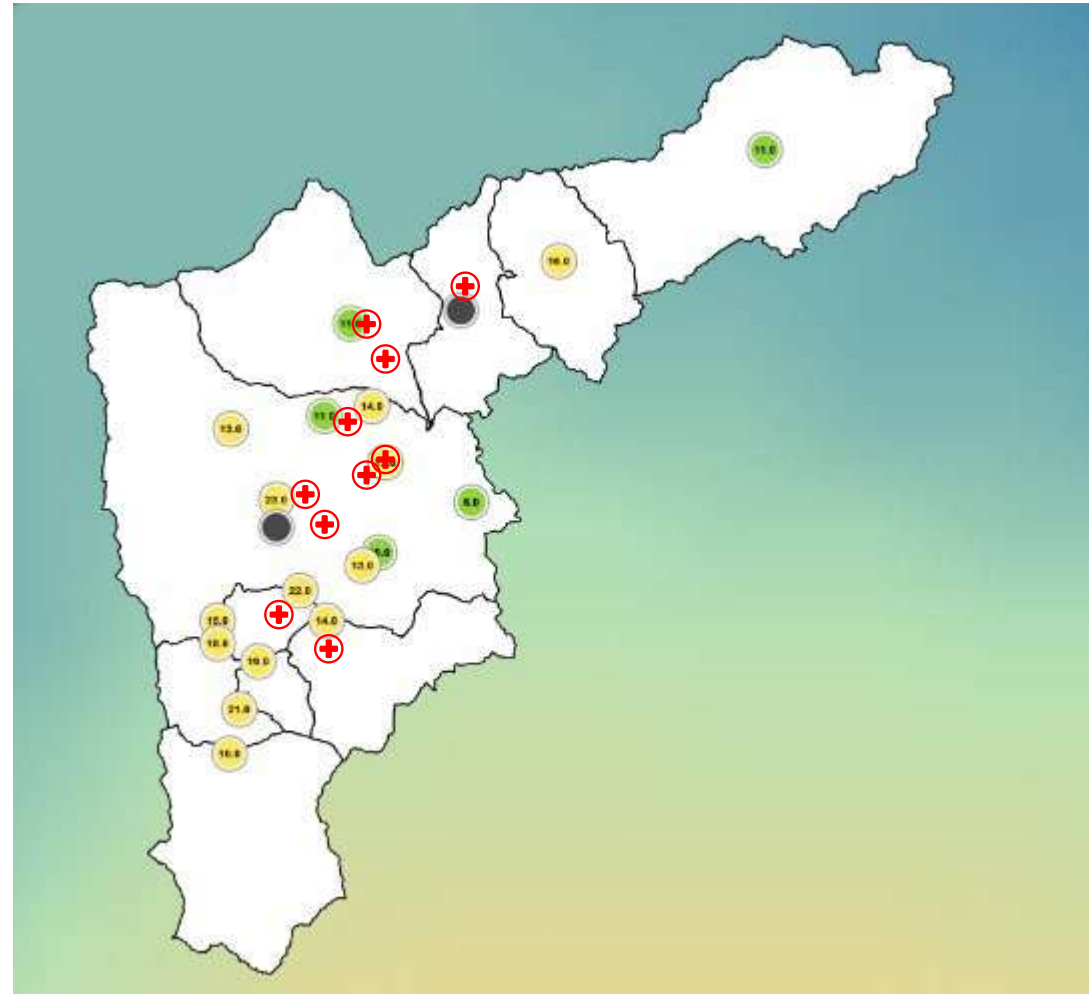
Datos distribuidos y heterogéneos , aportan generalización al modelo

Meteorológica y calidad del aire

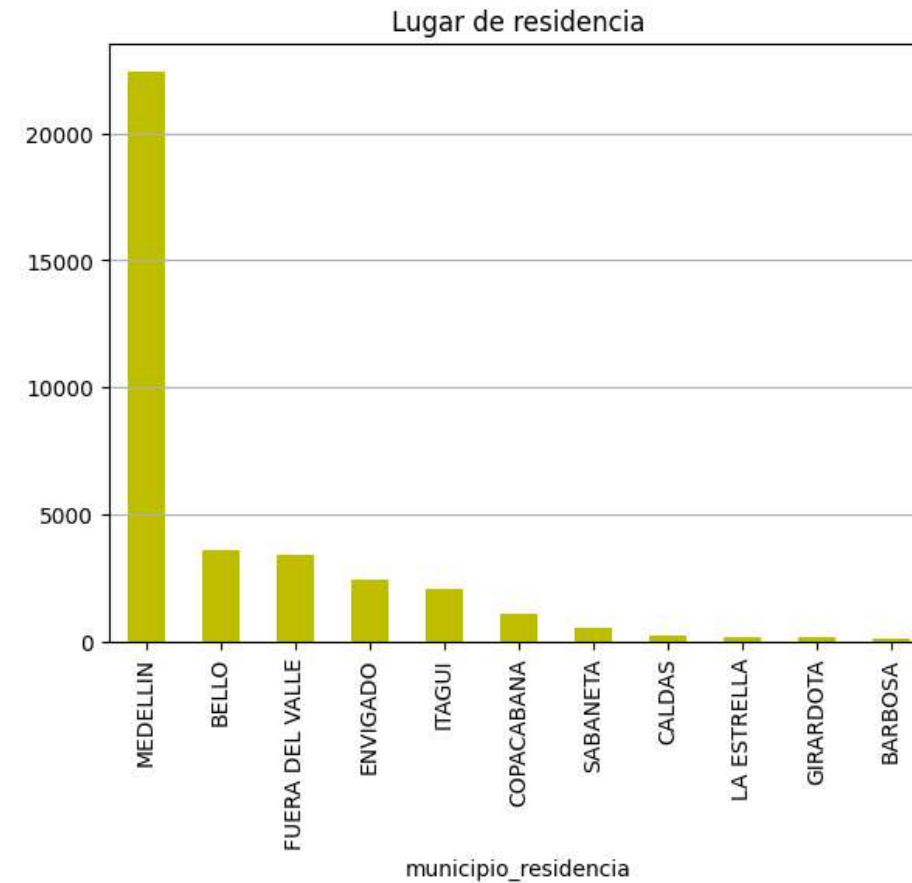
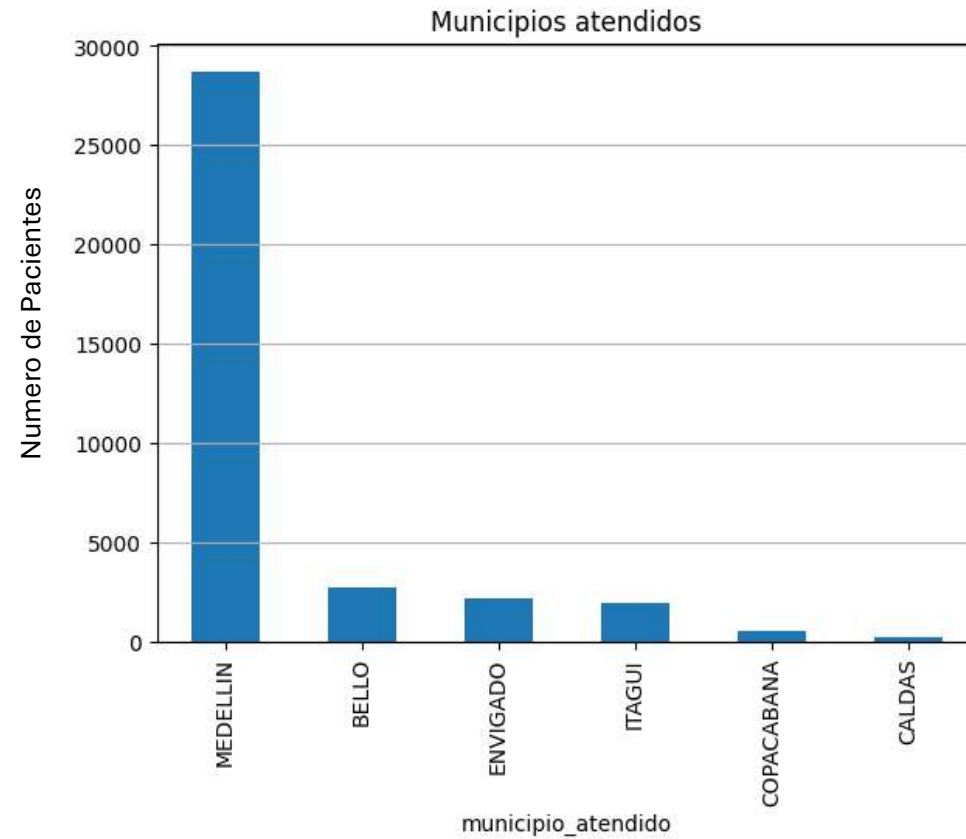
Sistema de Alerta temprana del Valle de Aburrá (SIATA)

Salud

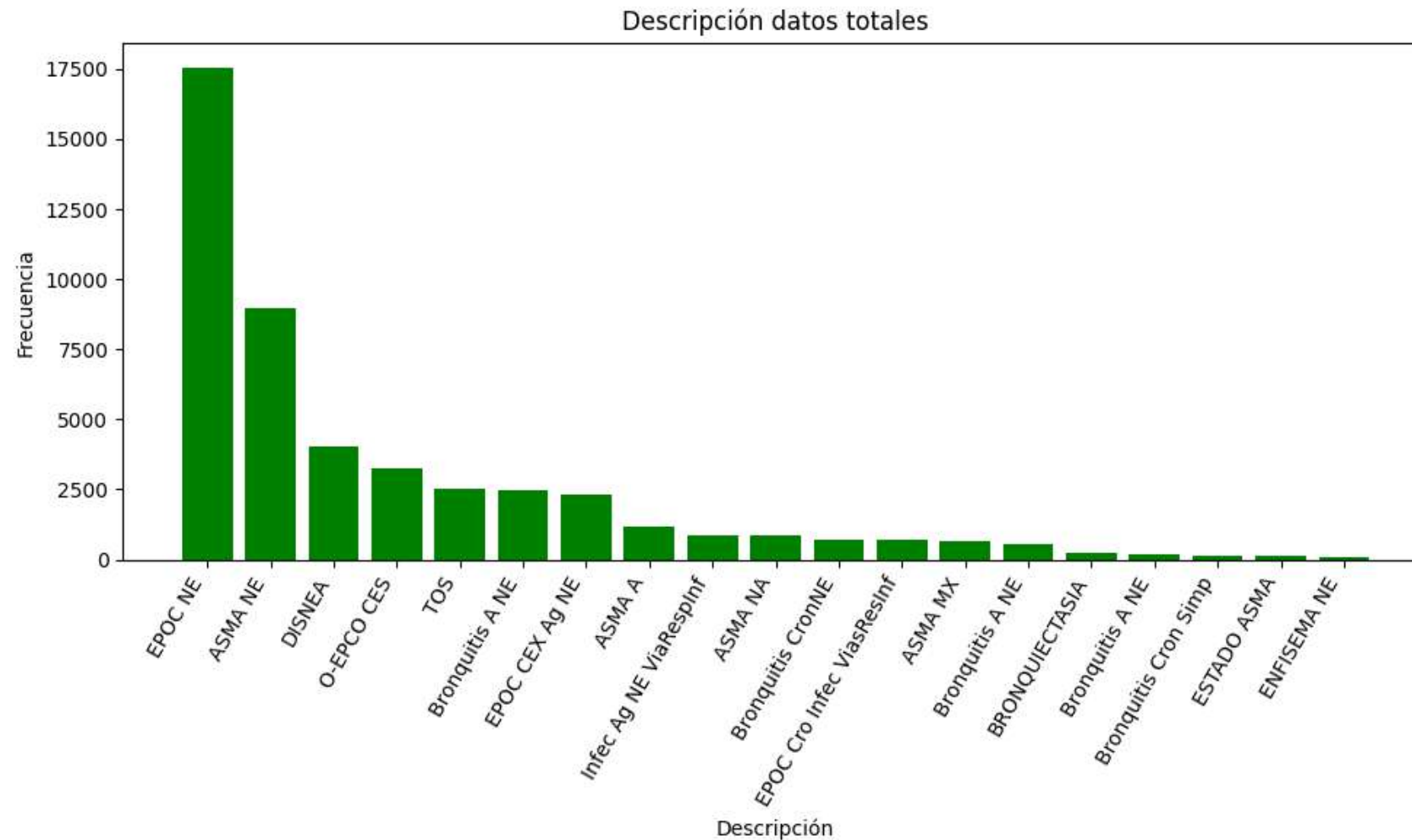
Institución médica Sumimedical SAS



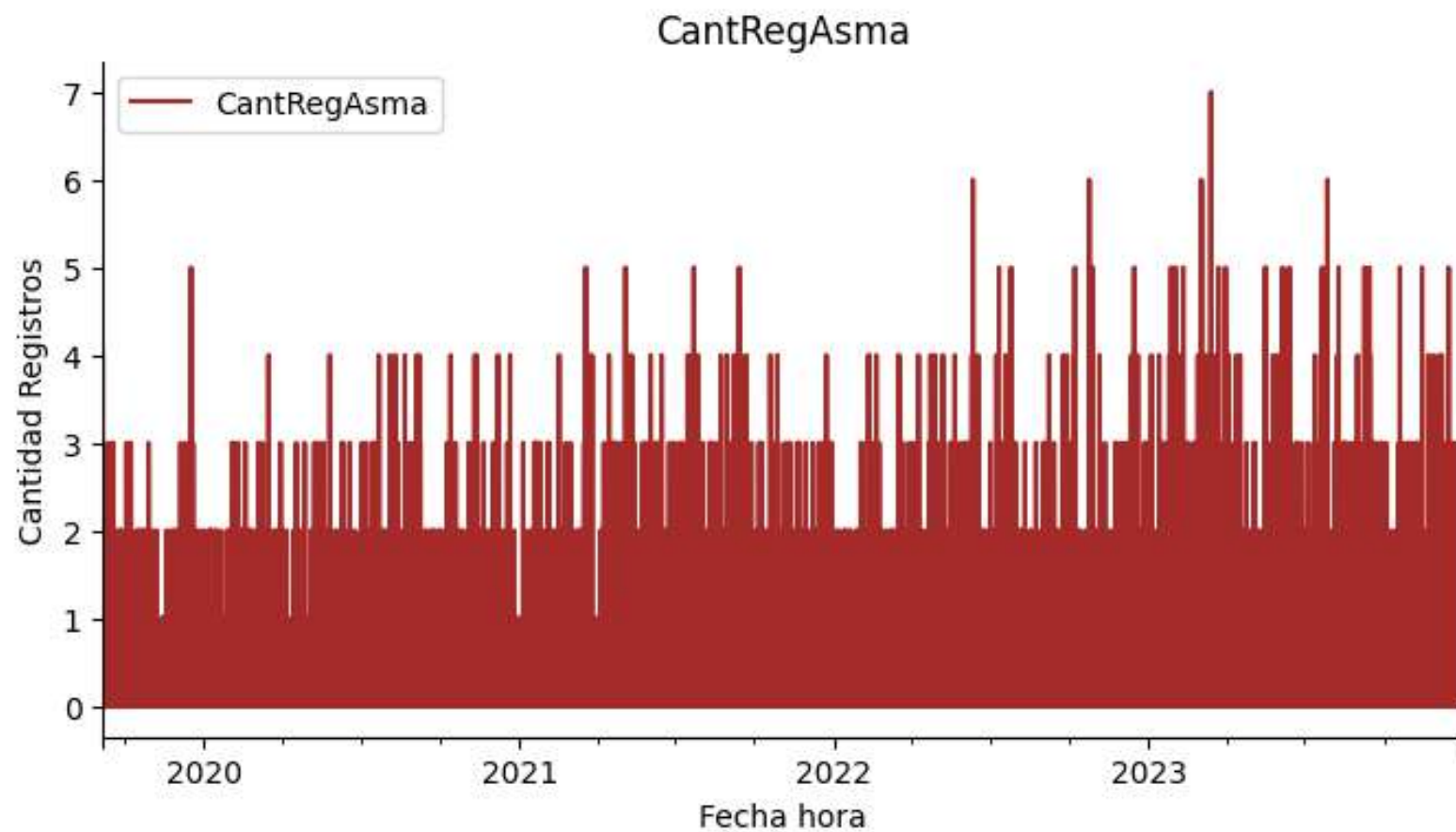
Resultados – Limpieza datos de Salud



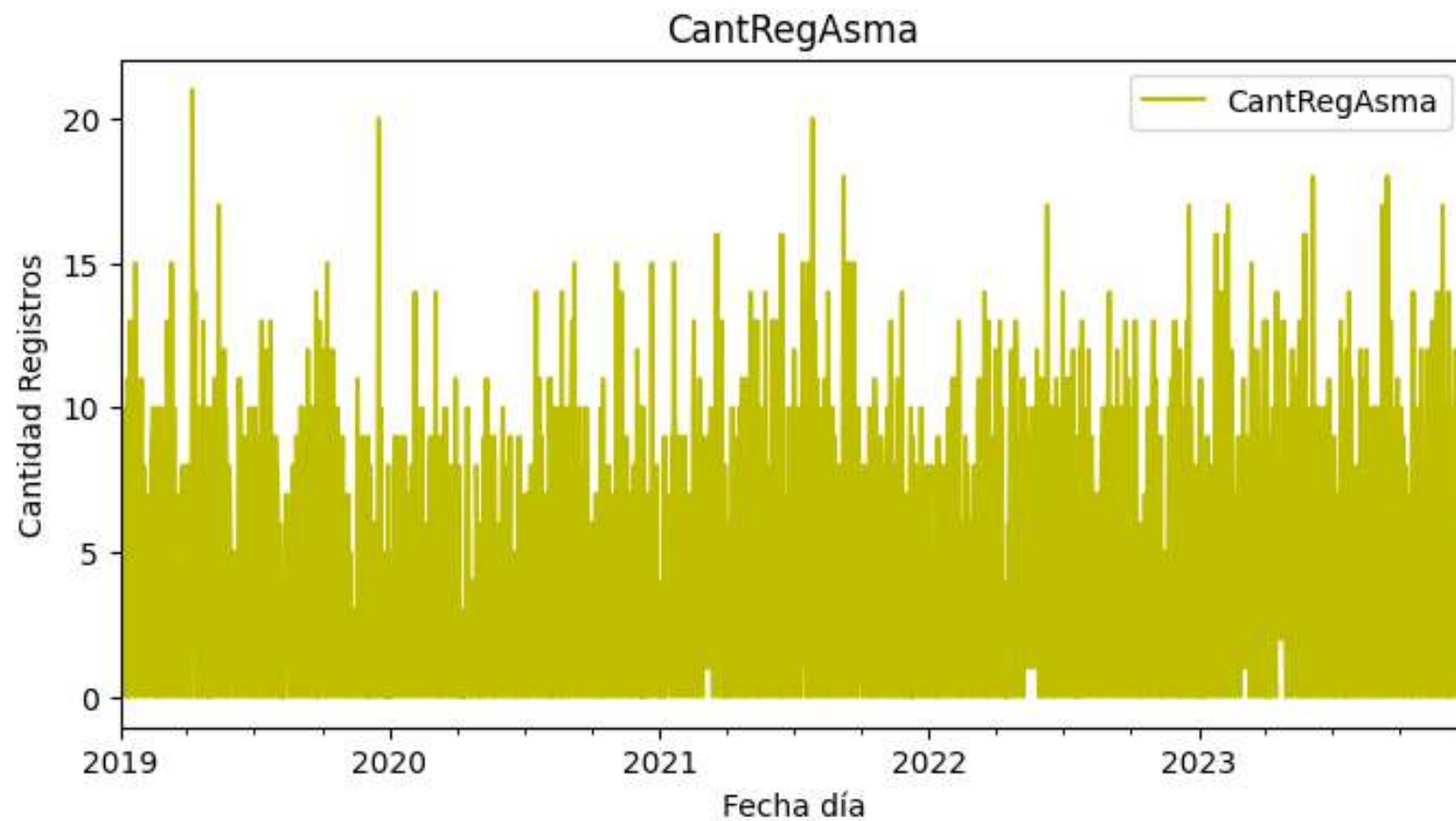
Resultados – Limpieza datos de Salud



Resultados – Datos de salud



Resultados – Datos de salud



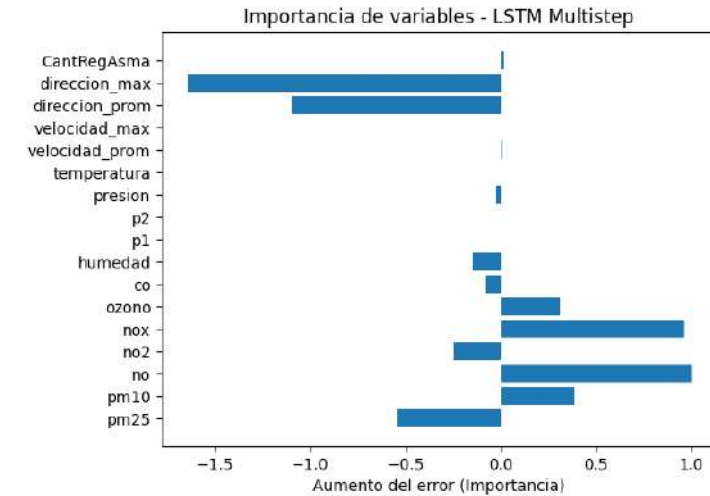
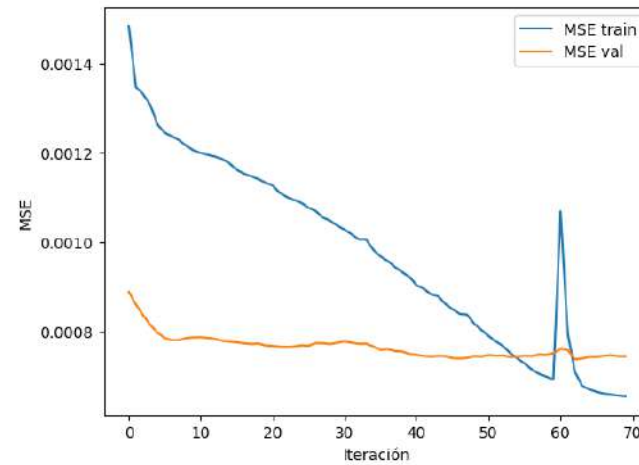
Resultados – Escenario 1

Rendimiento:

MSE train: 0.000659

MSE val: 0.000744

MSE test: 0.000740



Hiperparametros

horas entrenamiento = 240

horas predicción = 96

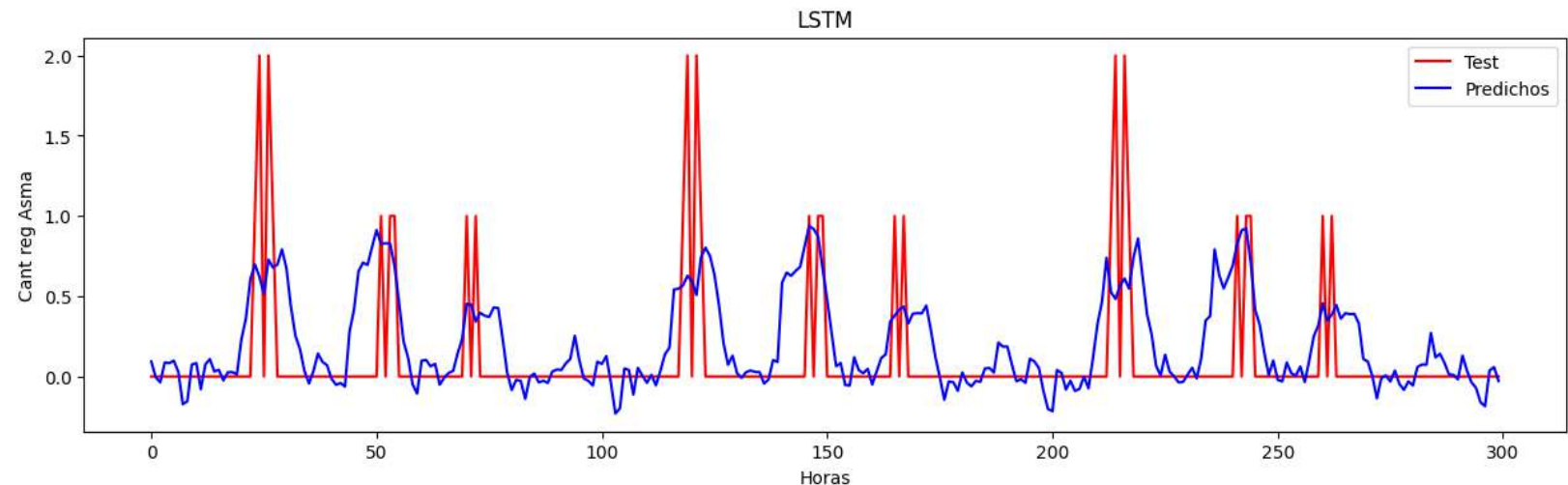
neuronas = 50

capas = 2 LSTM y 1 densa

Función de activación = linear

Épocas = 70

Tamaño de lote = 256



Resultados – Escenario 2

Rendimiento:

MSE train: 0.000560

MSE val: 0.000460

MSE test: 0.000463

Hiperparametros

horas entrenamiento = 480

horas predicción = 144

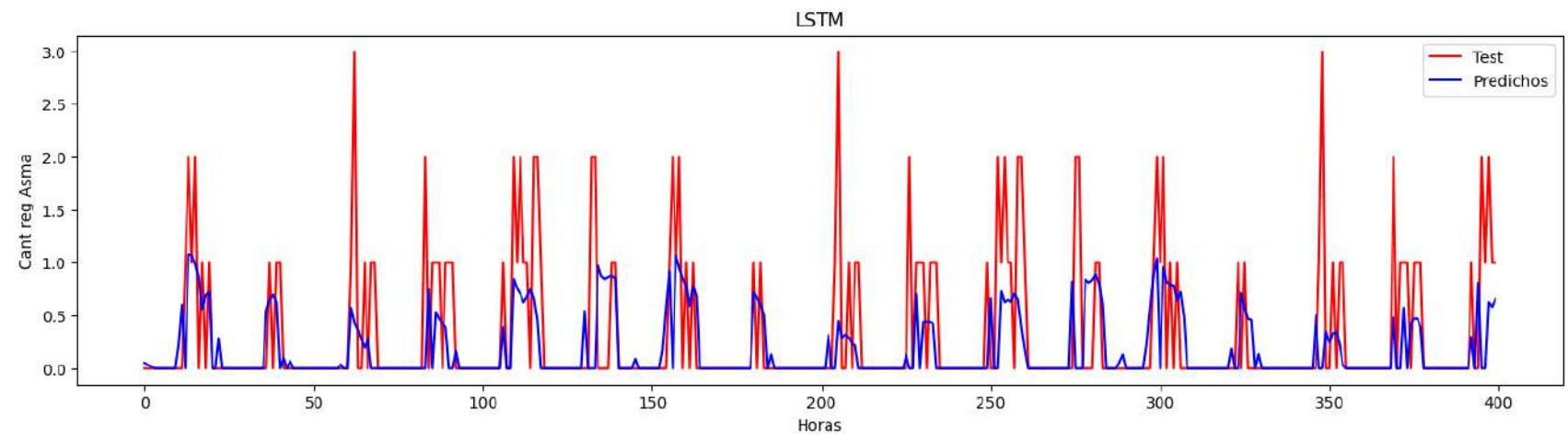
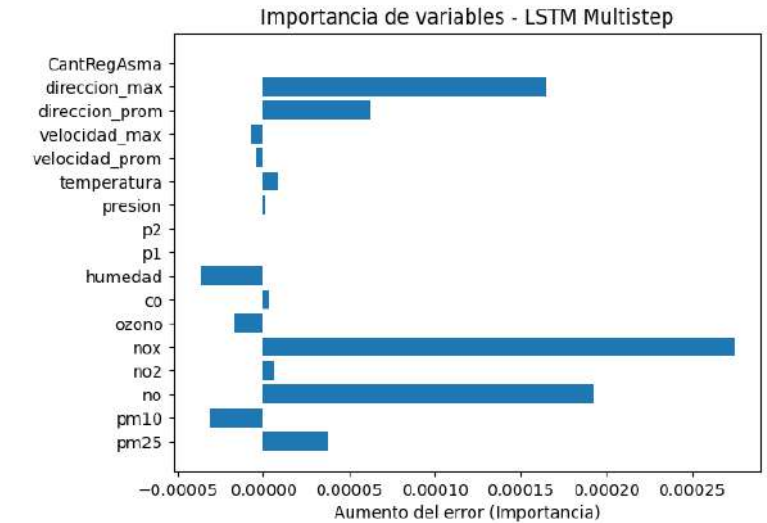
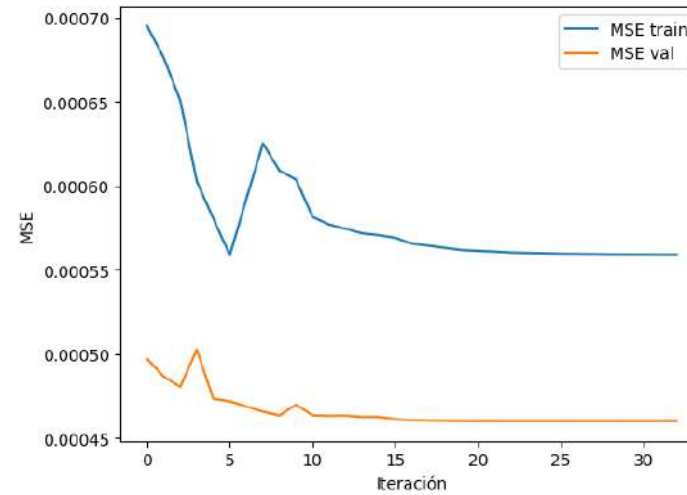
neuronas = 50

capas = 3 LSTM y 1 densa

Función de activación = relu

Épocas = 50

Tamaño de lote = 32



Conclusiones

- La **no linealidad** de este tipo de modelos, permite encontrar **patrones** que la estadística tradicional no haría
- La **fase de limpieza** de datos es **decisiva** para el desarrollo del modelo.
- Los **modelos** que permiten el desarrollo de **series temporales**, permiten un modelo **más robusto**.
- Los modelos **más utilizados son los LSMT**, sus variaciones y los modelos híbridos
- **La red convolucional de grafos y los modelos multivariantes espaciotemporales** tienen un **gran potencial** para la predicción relacionada con variables ambientales.

Trabajo Futuro

- Mejorar los métodos de **recopilación de datos de salud** para tener una bases de datos mas sólida
- **La implementación de mas tipos de modelos** mas avanzados como modelos que involucren la espacio temporalidad como la red neuronal de grafos convolucionales espaciotemporales, MST-GCN.
- Se hace necesario una **mayor validación externa** de los modelos de aprendizaje automático antes de que puedan implementarse de manera confiable en la práctica clínica real.
- **Mejorar la precisión y la aplicabilidad** de estos modelos mediante investigaciones y estudios de validación adicionales.

Agradecimientos



Bibliografía

- [1] Los costos en la salud asociados a la degradación ambiental en Colombia ascienden a \$20,7 billones. (2022). Dnp.gov.co.
[https://2022.dnp.gov.co/Paginas/Los-costos-en-la-salud-asociados-a-la-degradaci%C3%B3n-ambiental-en-Colombia-ascienden-a-\\$20,7-billones-.aspx](https://2022.dnp.gov.co/Paginas/Los-costos-en-la-salud-asociados-a-la-degradaci%C3%B3n-ambiental-en-Colombia-ascienden-a-$20,7-billones-.aspx)
- [2] Aguiar-Gil, D., Gómez-Peláez, L. M., Álvarez-Jaramillo, T., Correa-Ochoa, M. A., & Saldarriaga-Molina, J. C. (2020). Evaluating the impact of PM2.5 atmospheric pollution on population mortality in an urbanized valley in the American tropics. *Atmospheric Environment*, 224, 117343–117343.
<https://doi.org/10.1016/j.atmosenv.2020.117343>
- [3] De, L., Habitantes, L., Valle, D., & Aburrá, D. (2008). Y SUS EFECTOS SOBRE. https://www.metropol.gov.co/ambiental/calidad-del-aire/Biblioteca-aire/Estudios-calidad-del-aire/contaminacion_atmosferica_efectos_salud.pdf
- [4] Rosselli D., Gil Y., Lasalvia P., Peña M.P., Castañeda C., Virgüez O., González A., García C, Laucho Contreras M., Pontificia Universidad Javeriana, Neuroeconomix, Virrey Solís IPS, GlaxoSmithKline Colombia, Revista Colombiana de Neumología.

Más información



<https://casap.science/>



casap@casap.science

- [5] de, M. (2019). Aproximadamente 1 de cada 8 colombianos sufre asma. Retrieved April 28, 2025, from Minsalud.gov.co website: <https://www.minsalud.gov.co/Paginas/Aproximadamente-1-de-cada-8-colombianos-sufre-asma.aspx>
- [6] Orozco Levi M, Borrero L., Botero L., Londoño S., Saenz S., Servicio de Neumología , Fundación cardiovascular de Colombia, Bucaramanga Santander, Grupo de investigación EMICON . MINCINECIAS, Universidad de la Sabana, Sanofi Aventis de Colombia SA.
- [7] Aggarwal, C. C. (2023). Neural Networks and Deep Learning. In Neural Networks and Deep Learning. <https://doi.org/10.1007/978-3-031-29642-0>
- [8] Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. D. (2020). An Introduction to Machine Learning. Clinical Pharmacology and Therapeutics, 107(4). <https://doi.org/10.1002/cpt.1796>
- [9] Czum, J. M. (2020). Dive Into Deep Learning. Journal of the American College of Radiology, 17(5), 637–638. <https://doi.org/10.1016/j.jacr.2020.02.005>
- [10] Hunta, S., & Pengchata, R. (2022). A Deep Learning Approach for Prediction of Respiratory Disease from Air Quality. 6th International Conference on Information Technology, InCIT 2022, 262–265. <https://doi.org/10.1109/InCIT56086.2022.10067262>

Más información



<https://casap.science/>



casap@casap.science

- [11] Freeman, B. S., Taylor, G., Gharabaghi, B., & Thé, J. (2018). Forecasting air quality time series using deep learning. Journal of the Air and Waste Management Association, 68(8), 866–886. <https://doi.org/10.1080/10962247.2018.1459956>
- [12] Kim, M. S., Lee, J. H., Jang, Y. J., Lee, C. H., Choi, J. H., & Sung, T. E. (2020). Hybrid deep learning algorithm with open innovation perspective: A prediction model of asthmatic occurrence. Sustainability (Switzerland), 12(15). <https://doi.org/10.3390/su12156143>
- [13] Ram, S., Zhang, W., Williams, M., & Pengetnze, Y. (2015). Predicting Asthma-Related Emergency Department Visits Using Big Data. IEEE Journal of Biomedical and Health Informatics, 19(4), 1216–1223. <https://doi.org/10.1109/jbhi.2015.2404829>

Contactos: yesid.ramirez@udea.edu.co juan.lemos@udea.edu.co mandres.correa@udea.edu.co

Más información



<https://casap.science/>



casap@casap.science



Medellín, Colombia

CASAP X

CONGRESO COLOMBIANO Y
CONFERENCIA INTERNACIONAL

DE CALIDAD DE AIRE, CAMBIO CLIMÁTICO Y SALUD PÚBLICA



Más información



<https://casap.science/>



casap@casap.science