

**Modelo predictivo para la demanda mensual de
servicios de salud ocupacional 2025 de
la empresa SURA**

Luis Cabarcas Romero
Carlos Arango Mejía
Judith Pérez Conde

Universidad del Norte

Propuesta de reto
DATA CHALLENGE PRO
2025

Tabla de contenido

Resumen	3
1. Introducción	3
2. Planteamiento del problema	4
3. Metodología.....	5
3.1. Análisis de modelos predictivos	7
3.2. Justificación de la elección.....	9
4. Análisis de los datos.	10
5. Desarrollo del modelo predictivo	11
6. Resultados y visualizaciones	11
7. Conclusiones.....	13
8. Referencias	14

Resumen

Este informe presenta el desarrollo de un modelo predictivo para estimar la demanda mensual de servicios de salud ocupacional en Colombia durante el año 2025, con énfasis en atenciones por accidentes y enfermedades laborales. La propuesta fue diseñada para la empresa ARL SURA, en respuesta a la necesidad de mejorar la planificación y distribución de recursos en sus distintos municipios de operación.

Se empleó una metodología basada en ciencia de datos, que incluyó la recolección, limpieza, transformación y análisis exploratorio de más de 11 millones de registros históricos. A partir de estos datos, se evaluaron tres modelos predictivos de series temporales: LSTM, ARIMA+XGBoost y N-BEATS. El modelo seleccionado fue N-BEATS, debido a su superior desempeño en métricas de error como RMSE (27.74), MAE (19.36).

Los resultados predicen tendencias diferenciadas según el municipio. Por ejemplo, mientras que Medellín presenta una reducción progresiva en la demanda, Barranquilla muestra un crecimiento exponencial, lo cual evidencia la importancia de generar predicciones localizadas. Este trabajo proporciona una herramienta útil para la toma de decisiones estratégicas en salud ocupacional, permitiendo anticipar necesidades y optimizar la respuesta ante la variabilidad de la demanda.

1. INTRODUCCIÓN

En el entorno actual, donde las organizaciones enfrentan constantes desafíos en la gestión de riesgos laborales, la atención oportuna y eficiente de accidentes y enfermedades asociadas al trabajo cobra una relevancia cada vez mayor. Las Administradoras de Riesgos Laborales (ARL), como SURA, deben garantizar una adecuada prestación de servicios de salud para sus afiliados, lo que implica no solo reaccionar ante los eventos, sino también anticiparse a ellos mediante estrategias de planificación y análisis predictivo. Así pues, la investigación sobre la demanda futura de servicios de salud laboral pone de relieve varias tendencias clave. Los modelos de simulación pueden ayudar a predecir la demanda y, en consecuencia, a optimizar la utilización de recursos para enfermedades de alto coste (Walter Toro Jiménez, 2005). Esto se traduce en no solo una mejora de eficiencia operativa, sino que también marca una diferencia en términos de vidas salvadas.

Entonces, el reto planteado surge de esta necesidad: desarrollar un modelo que permita predecir la demanda de diferentes servicios médicos derivados de incidentes laborales en distintos municipios del país.

La predicción de la demanda de servicios sanitarios es crucial para una gestión eficiente de los recursos. Por ejemplo, Ramos et al. (2019) propusieron un sistema experto que utiliza algoritmos de minería de datos para analizar los factores que influyen en los patrones de llamadas en un sistema de gestión de relaciones con los clientes, con el objetivo de predecir la variabilidad de la demanda de asistencia

sanitaria. Álvarez García (2011) examinó modelos de demanda sanitaria en España, investigando los factores que afectan a las consultas médicas y a la utilización de la atención dental. Ramírez-Tinoco et al. (2020). Estos estudios ponen de relieve la importancia de los modelos predictivos en la planificación de la asistencia sanitaria.

Buscar una solución a esta problemática no solo es pertinente, sino fundamental, ya que una predicción precisa de la demanda permite una mejor asignación de recursos, reduce tiempos de respuesta, mejora la cobertura y calidad del servicio, y optimiza la operación de los sistemas de salud ocupacional. En última instancia, este tipo de herramientas analíticas contribuyen a preservar la salud y seguridad de los trabajadores, al tiempo que fortalecen la eficiencia de las instituciones responsables de su cuidado.

2. PLANTEAMIENTO DEL PROBLEMA

La prestación de servicios de salud ocupacional por parte de las Administradoras de Riesgos Laborales (ARL) enfrenta incertidumbres en la demanda de atenciones por accidentes y enfermedades laborales, lo que dificulta la asignación oportuna de recursos en cada municipio. Sin embargo, las herramientas actuales rara vez integran de forma simultánea patrones de estacionalidad, tendencias históricas, variables demográficas e indicadores económicos locales, elementos esenciales para capturar la complejidad real de la demanda.

La falta de un sistema predictivo robusto limita la capacidad de ARL SURA para planificar con antelación la dotación de personal, insumos y equipos médicos en los distintos municipios, lo cual puede derivar en demoras en la atención y en la subutilización o desbordamiento de recursos. Además, la variabilidad geográfica y temporal de los siniestros laborales exige un modelo que ofrezca predicciones por municipio y tipo de servicio médico, incluyendo específicamente atenciones por accidentes y enfermedades laborales.

Este estudio presenta limitaciones derivadas de la calidad y cobertura de los datos históricos, donde registros faltantes o inconsistentes pueden introducir sesgos en el análisis. Asimismo, la heterogeneidad de los factores externos —como las diferencias demográficas y económicas entre municipios— puede no estar completamente reflejada en el conjunto de datos, lo que afecta la validez de las predicciones. La demanda de servicios de salud también exhibe patrones de estacionalidad y picos atípicos durante eventos como campañas de seguridad o crisis sanitarias, así como variaciones relacionadas con fines de semana y feriados que modifican el comportamiento de la serie temporal. (Shaw et al., 2022)

Pregunta problema: ¿Cómo predecir, mediante un modelo univariado de series de tiempo la demanda mensual de la cantidad de servicios de salud ocupacional en énfasis de enfermedades laborales en cada municipio de Colombia?

3. METODOLOGÍA

La metodología aplicada en esta investigación se fundamentó en un enfoque cuantitativo de análisis de datos, mediante técnicas de Ciencia de Datos orientadas al análisis exploratorio y la preparación de datos de salud. Se utilizó Python como lenguaje de programación para desarrollar y dar solución al problema. A continuación, se describen las etapas desarrolladas durante el estudio:

▪ Importación de librerías y datos

Se inició con la importación de las principales librerías de análisis y visualización de datos en Python, tales como pandas, numpy, matplotlib, seaborn, y statsmodels. Posteriormente, se cargó el conjunto de datos en formato .csv que contiene información relacionada con conceptos de facturación en salud, diagnósticos CIE-10, fechas de atención, municipios, cantidades y tipificaciones, entre otras variables.

▪ Análisis Exploratorio de Datos (EDA)

Con el fin de conocer la estructura, calidad y distribución de los datos, se llevó a cabo un análisis exploratorio, el cual consistió en un primer acercamiento a la estructura del conjunto de datos. Se identificaron el número de registros, tipos de variables, valores únicos, datos faltantes y estadísticas descriptivas. Este paso permitió comprender la naturaleza y complejidad de la información disponible, así como detectar posibles inconsistencias o valores atípicos.

▪ Visualización de datos

Para facilitar la interpretación de los datos, se realizaron diversas visualizaciones, incluyendo: gráficos de barras para observar la frecuencia de códigos CIE-10, conceptos de factura y tipificaciones; gráficos de líneas para analizar tendencias temporales en los registros; boxplots para examinar la distribución y detectar valores atípicos en las cantidades registradas; diagramas de descomposición estacional en series temporales, utilizando el modelo STL (Seasonal-Trend decomposition using Loess), con el fin de identificar patrones estacionales, tendencias y componentes residuales.

Por ejemplo, en la figura 1, se presenta un gráfico de barras que muestra los diez diagnósticos CIE-10 más frecuentes en el conjunto de datos. Esta visualización permitió identificar rápidamente los códigos más recurrentes, útiles para posteriores análisis de concentración diagnóstica y consumo de recursos.

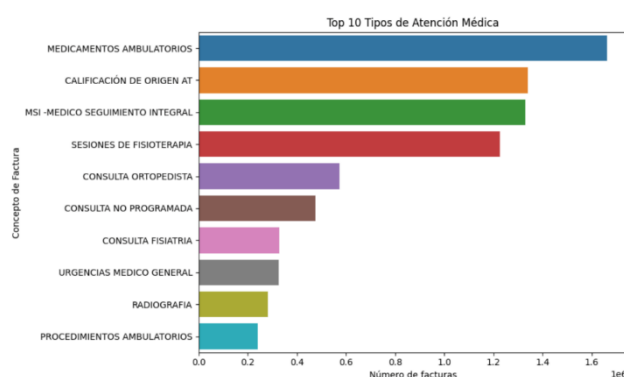


Figura 1: Tipos de atención médica

Asimismo, la figura 2 presenta una descomposición estacional de la serie temporal de registros de atención, lo que permitió evidenciar una tendencia creciente en ciertos meses del año, posiblemente relacionada con campañas de salud pública o ciclos epidemiológicos.

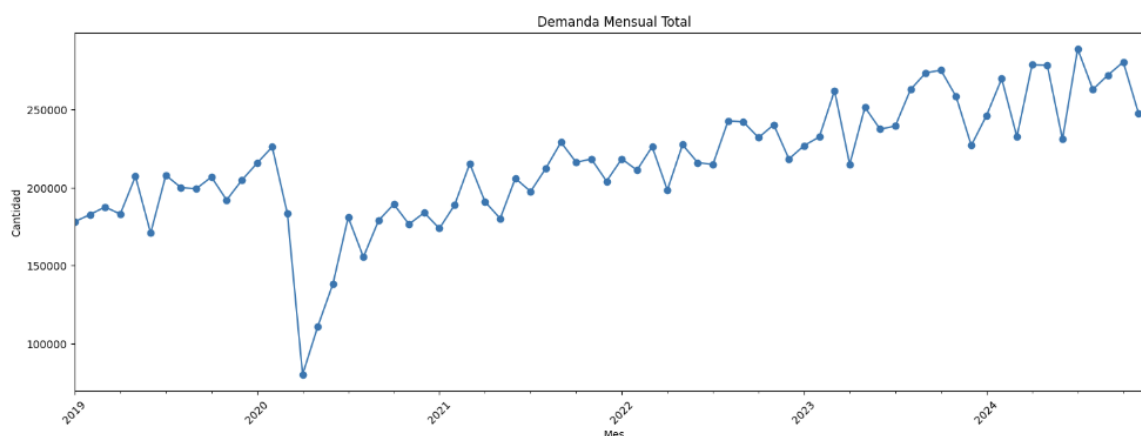


Figura 2: Registros de atención por mes

En adición, la figura 3 presenta un gráfico de barras con los municipios con mayor demanda de servicios, estando Bogotá, Medellín y Santiago de Cali en el top 3.

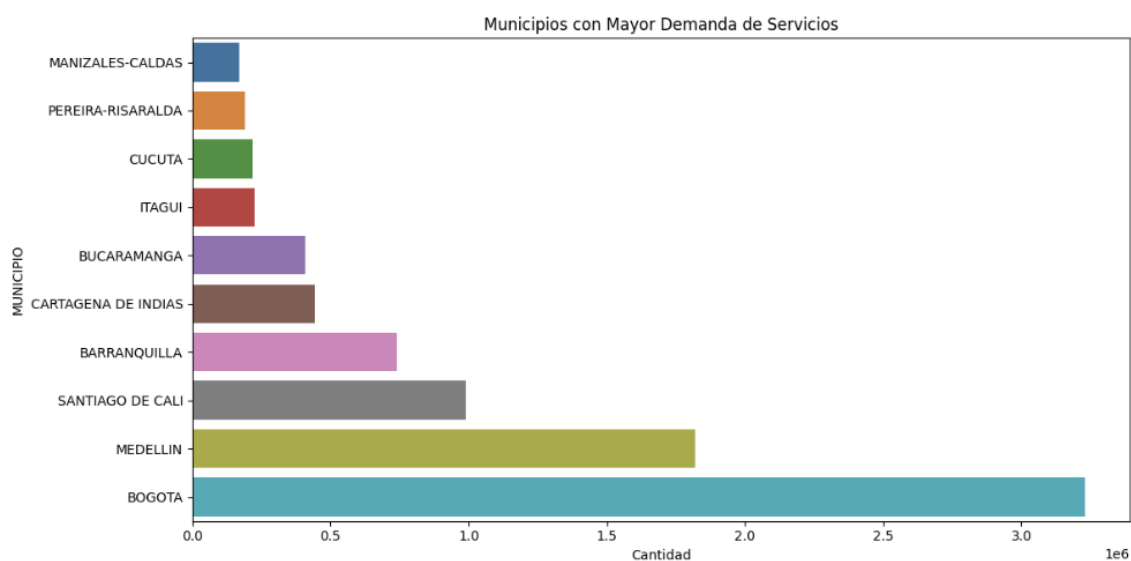


Figura 3: Municipios con mayor demanda de servicios

▪ Limpieza de datos

En este apartado se procedió con la depuración de los datos, corrigiendo errores, eliminando valores duplicados y gestionando los registros incompletos o nulos. Esta etapa fue fundamental para garantizar la calidad del dataset y la fiabilidad de los resultados posteriores. Se emplearon técnicas como la imputación de datos nulos, normalización de formatos de fechas y eliminación de registros inconsistentes. Asimismo, se consolidaron las variables categóricas y se eliminaron columnas irrelevantes para el análisis.

Por ejemplo, en la tabla 1 se resume algunas de las acciones realizadas durante la limpieza, con el objetivo de dejar trazabilidad de las decisiones tomadas en esta etapa crítica del análisis.

Tabla 1: Acciones de limpieza de datos

Acción realizada	Descripción breve	Nº de registros afectados
Eliminación de duplicados	Se eliminaron filas completamente repetidas	220152
Conversión de tipos de datos	Fechas transformadas a tipo datetime	100% del dataset
Normalización de conceptos de factura	Se homogeneizaron nombres con errores tipográficos	-

Finalmente, con base en lo anterior, se construyeron nuevas variables para tener un mejor análisis del comportamiento de los datos. Entre estas se encuentran variables temporales derivadas de las fechas (mes, día, año), categorización de los conceptos de factura, codificación de variables categóricas, y generación de indicadores que reflejan comportamientos atípicos o inusuales. Esto se hizo sobre la base teórica del Feature Engineering (La ingeniería de características), la cual es un proceso crucial en el aprendizaje automático que se centra en mejorar la calidad de los datos y el rendimiento de los modelos mediante la creación, transformación y selección de características informativas (Verdonck et al., 2021; Soviany & Soviany, 2020).

3.1. Análisis de modelos predictivos

Para aprovechar al máximo los registros históricos de cantidad de servicios por municipio, tipo de servicio, variables demográficas y económicas, cada modelo requiere un conjunto de pasos de preparación y ajuste específicos. LSTM captura

dependencias a largo plazo al incorporar regresores exógenos directamente en la arquitectura recurrente. N-BEATS ofrece un enfoque “end-to-end” que aprende automáticamente tendencias y estacionalidades sin ingeniería manual. El híbrido ARIMA + XGBoost combina la claridad de los componentes lineales con la potencia de los árboles para modelar residuos y no linealidades. Cada uno aporta ventajas distintas para optimizar la precisión y la interpretabilidad en el entorno de ARL SURA.

Se consideró el uso de diferentes modelos predictivos; en esta sección se presenta un resumen de las características de cada uno.

LSTM (Long Short-Term Memory)

Las redes LSTM (Long Short-Term Memory) son recurrentes y están diseñadas para capturar dependencias a largo plazo en series temporales univariadas y multivariadas. Para su correcto funcionamiento, es esencial normalizar la serie y las variables exógenas (por ejemplo, demográficas y económicas) mediante técnicas como MinMax o Z-score, lo que mejora la convergencia durante el entrenamiento. La creación de ventanas deslizantes (“look-back”) transforma la serie en muestras de forma (samples, timesteps, features), permitiendo que la LSTM aprenda patrones temporales complejos. Es necesario aplicar regularización (Dropout) y “early stopping” para evitar el sobre ajuste, así como validar mediante cross-validation con ventanas móviles que respeten el orden temporal (Stack Exchange, 2022).

N-BEATS

N-BEATS es una arquitectura de deep learning específica para forecasting univariado, basada en bloques residuales que modelan tendencias y estacionalidades sin ingeniería manual de características. Su diseño fully-connected facilita un entrenamiento rápido y paralelizable, convergiendo en pocas épocas incluso con series largas. La variante interpretable impone restricciones en las funciones base (polinómicas y de Fourier), lo que ofrece explicabilidad sobre cómo se componen tendencia y estacionalidad. N-BEATS(P) permite escalar el entrenamiento a múltiples series simultáneamente, manteniendo eficiencia computacional (Oreshkin, Dudek, Peřka, & Turkina, 2020).

ARIMA + XGBoost

El enfoque híbrido combina un modelo ARIMA para extraer componentes lineales de tendencia y estacionalidad con XGBoost para modelar residuales y relaciones no lineales. Primero se ajusta un ARIMA/SARIMAX, determinando órdenes (p,d,q) con ACF/PACF, y se obtienen los residuales. Luego, XGBoost se entrena sobre esos residuales incorporando lags, medias móviles y variables exógenas, optimizando hiperparámetros (n_estimators, max_depth, learning_rate) vía grid search. Finalmente,

la predicción total resulta de sumar las predicciones de ARIMA y XGBoost, obteniendo un balance entre interpretabilidad y precisión (Odhiambo et al., 2024).

La consideración de estos modelos predictivos se debe al análisis de los modelos más demandados y usados en la industria de la salud tales como las aseguradoras de salud como Cigna Healthcare han integrado modelos híbridos ARIMA–XGBoost para anticipar la carga de trabajo en sus unidades de atención por problemas musculoesqueléticos. Este enfoque permitió reducir el error porcentual medio (MAPE) de alrededor de 25 % (con ARIMA puro) a menos de 12 % tras incorporar XGBoost sobre los residuales, mejorando la asignación de citas y la programación de especialistas (Odhiambo, Cornelious & Waititu, 2024). De igual manera el caso de la empresa en el sector público, el Servicio Nacional de Salud del Reino Unido (NHS) aplica modelos SARIMA y Prophet–LSTM para pronosticar las admisiones en urgencias (A&E). Durante la pandemia de COVID-19, integraron variables de movilidad y de incidencia viral como regresores en un Prophet–LSTM híbrido, logrando predecir picos de demanda con un horizonte de tres semanas y reduciendo los desbordamientos en sala de espera en un 20 % (Taylor & Letham, 2018; Quinn et al., 2024). Nos dieron indicios que estos serían los modelos más convenientes para el desarrollo del modelo predictivo.

3.2. Justificación de la elección

Se implementaron las alternativas planteadas en el apartado anterior tras la evaluación de los tres modelos considerados, el modelo N-BEATS es el más adecuado para estimar la demanda mensual de servicios de salud ocupacional. Esta decisión se fundamentó en el rendimiento superior del modelo frente a otras alternativas (véase tabla 2). Específicamente, N-BEATS obtuvo un RMSE de 27.74 y un MAE de 19.36, demostrando una mayor precisión y menor error en las predicciones. En comparación, el modelo LSTM presentó un RMSE de 60.79 y un MAE de 49.95, mientras que el modelo combinado (ARIMA + XGBoost) alcanzó un RMSE de 36.78 y un MAE de 29.98.

GPUs detectadas: 2

Procesando 10 combos

Tabla 2: métricas por modelo

	RMSE	MAE	MODEL
0	36.778584	29.976001	Combined
1	60.789555	49.947600	LSTM
2	27.739709	19.356332	NBEATS

Mejor modelo: NBeats

Pipeline completado.

Por otro lado, es importante tener en cuenta que la variable objetivo Cantidad corresponde a la suma diaria de todos los servicios registrados para cada combinación

de municipio y tipo de atención. Esto implica que los valores de dicha variable se encuentran en una escala agregada, con magnitudes que varían entre decenas y cientos de registros diarios según el contexto geográfico y el tipo de servicio.

Como resultado, las métricas de error absoluto como el RMSE (Root Mean Squared Error) y el MAE (Mean Absolute Error) están directamente influenciadas por esta escala. Por ejemplo, un RMSE de 36 puede ser completamente razonable si la demanda diaria promedio es de 300 servicios, pero podría representar un error importante si la media es cercana a 50.

Por esta razón, estas métricas deben interpretarse en relación con la magnitud de la demanda promedio para cada serie temporal. En este sentido, el uso de indicadores relativos como el SMAPE (Symmetric Mean Absolute Percentage Error) y el MASE (Mean Absolute Scaled Error) se vuelve especialmente útil, ya que permiten comparar el rendimiento del modelo entre combinaciones con diferentes volúmenes de servicio.

Ahora bien, retomando lo N-BEATS, además de su desempeño cuantitativo, ofrece ventajas técnicas que lo hacen especialmente adecuado para este tipo de series temporales univariadas. Su arquitectura basada en bloques residuales completamente conectados permite modelar tendencias y estacionalidades sin requerir ingeniería de características, facilitando una implementación más sencilla y explicable. Esta capacidad para ofrecer interpretabilidad es particularmente valiosa en contextos donde se requiere justificar las decisiones predictivas ante entidades técnicas o institucionales (Oreshkin, Carпов, Chapados & Bengio, 2019).

4. ANÁLISIS DE LOS DATOS.

Obtención de los datos

Los datos son proporcionados por la empresa ARL SURA donde nos ofrecen 3 archivos para el desarrollo del modelo predictivo de la demanda de servicios de salud ocupacional. Estos archivos contienen información detallada sobre atenciones médicas relacionadas con accidentes y enfermedades laborales, distribuidas por municipio y tipo de servicio.

1. Información Salud 2019-2024.txt

Este conjunto de datos constituye la base principal para el análisis y modelado. Incluye registros individuales de atenciones médicas, con campos que abarcan desde la identificación de la oficina de ARL hasta detalles específicos de la atención médica. Teniendo en cuenta que algunos de los campos más relevantes son: Nombre_Tipo_Atencion_Arp, Categoria_CIE10, Siniestro_Diagnosti_Princi_Id, FECHA_ATENCION, Nombre_Municipio_IPS, Geogra_Municipio_Id, Nombre_Departamento_IPS, entre otros. Los cuales se emplearon para desarrollar el modelo predictivo propuesto en el informe.

2. Red Prestadores.xlsx

El presente archivo ofrece un resumen consolidado de la demanda máxima observada para distintos servicios médicos en cada municipio, identificados mediante el campo Geogra_Municipio_Id. Incluye variables como Concepto_Factura_Desc, que describe el tipo de servicio prestado, y max_cantidad, que indica la mayor cantidad registrada de dicho servicio.

Además, el campo HOMOLOGACION NIT permite asociar los datos con entidades específicas. Este conjunto de datos identificar los servicios más demandados por municipio y facilitar la priorización en la asignación de recursos.

3. Diccionario de datos UNINORTE.xlsx

Este archivo nos brinda una información resumida del principal, donde detalla las autorizaciones médicas derivadas de siniestros laborales, incluye variables como la descripción del concepto de facturación, el identificador geográfico del municipio (Geogra_Municipio_Id), la cantidad máxima registrada (max_cantidad) y la homologación del NIT de la IPS o establecimiento de salud. Estos datos permiten analizar las variables que influyen en el conjunto de datos proporcionado, su distribución y la frecuencia de los servicios médicos autorizados, facilitando el estudio de patrones en la atención médica relacionada con siniestros laborales.

5. DESARROLLO DEL MODELO PREDICTIVO

Con base en los resultados obtenidos durante el análisis exploratorio, se procedió a la implementación de diversos modelos de predicción de series temporales con el objetivo de estimar la demanda futura de servicios de salud en distintos municipios. Para ello, se seleccionaron combinaciones representativas de municipio y servicio, y se desarrolló un pipeline de modelado que contempla tres enfoques principales: modelos estadísticos, modelos de aprendizaje automático y modelos basados en redes neuronales.

Se trabajó con un conjunto de datos en formato panel, reorganizado para generar series temporales univariadas por cada combinación de municipio y servicio. Estas series fueron normalizadas y divididas en conjuntos de entrenamiento y prueba, manteniendo su estructura temporal. El horizonte de predicción se estableció considerando el contexto del problema y la estacionalidad identificada previamente.

6. RESULTADOS Y VISUALIZACIONES

A manera de ejemplo, se muestran los resultados del modelo para las cinco ciudades con mayor demanda de servicios médicos (identificadas en el análisis exploratorio), específicamente en el requerimiento de medicamentos ambulatorios.

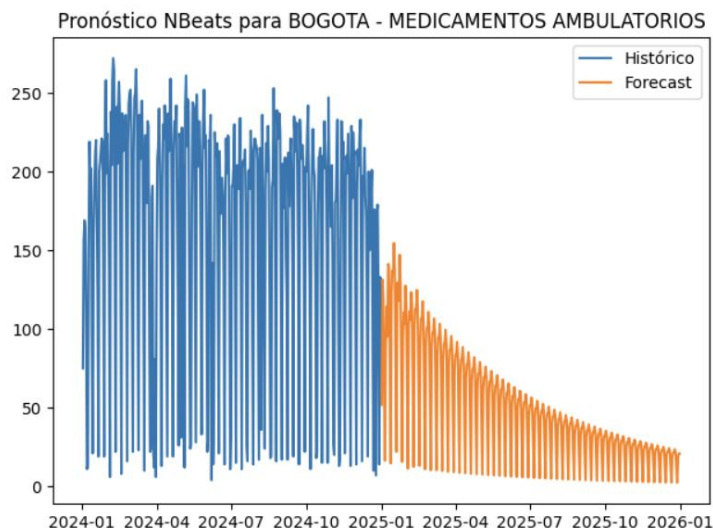


Figura 4: Pronóstico NBeats Bogotá - medicamentos ambulatorios

La figura 4 muestra la predicción de demanda de medicamentos ambulatorios en la ciudad de Bogotá. Según el modelo, estos requerimientos tienen tendencia a disminuir. En las figuras 5, 6 y 7 se observa un comportamiento similar, correspondiendo a las ciudades de Medellín, Santiago de Cali y Cartagena de Indias, respectivamente.

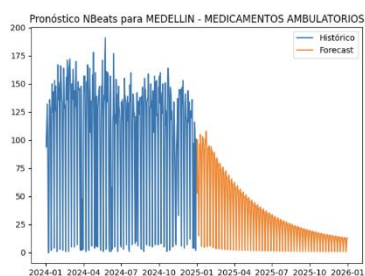


Figura 5: Pronóstico NBeats Medellín - Medicamentos ambulatorios

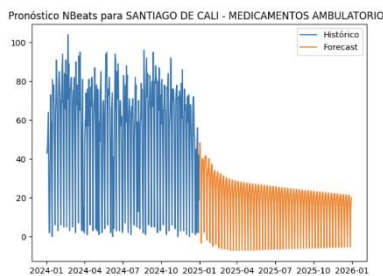


Figura 6: Pronóstico NBeats Cali - Medicamentos ambulatorios

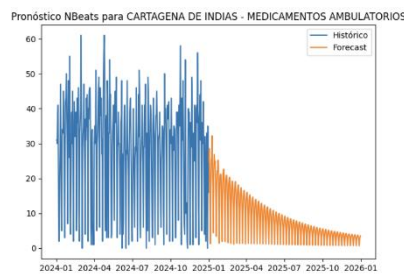


Figura 7: Pronóstico NBeats Cartagena - Medicamentos ambulatorios

En cuanto a Barranquilla, durante el período histórico (2024), la demanda diaria del municipio de Barranquilla se mantuvo oscilando entre 40 y 90 unidades, con una variabilidad constante, pero sin una tendencia marcada. No obstante, a partir de inicios del año 2025, se evidencia un cambio significativo en la tendencia: el modelo proyecta un aumento sostenido y exponencial en el consumo, alcanzando valores superiores a 130 unidades diarias hacia finales del mismo año (véase la figura 8).

Este comportamiento pronosticado podría responder a factores como un crecimiento poblacional, mayor cobertura del sistema de salud, o cambios en la prescripción de tratamientos ambulatorios. La pendiente positiva sugiere que el sistema deberá prepararse para un incremento considerable en la demanda. Tomando como referencia el segundo semestre de 2025, donde el promedio diario estimado ronda las 100

unidades, se puede proyectar una demanda mensual cercana a 3.000 unidades, lo que representa un crecimiento notable frente al histórico. Este análisis es clave para la planificación logística y presupuestal en la cadena de suministro de medicamentos en la región.

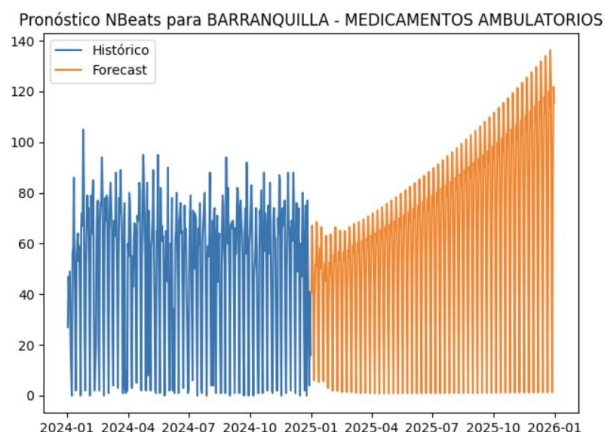


Figura 8: Pronóstico NBeats Barranquilla - Medicamentos ambulatorios

7. CONCLUSIONES

El desarrollo de este modelo predictivo permitió anticipar la demanda mensual de servicios de salud ocupacional a nivel municipal. A través de la estructuración y análisis de series temporales, se identificaron patrones históricos de consumo que facilitaron la formulación de predicciones útiles para la planificación de recursos en atención médica por accidentes y enfermedades laborales. El modelo tiene potencial de ser integrado en los procesos operativos de asignación y cobertura de servicios de las administradoras de riesgos laborales.

Durante la fase de modelado, se implementaron y compararon tres enfoques: LSTM, ARIMA+XGBoost y N-BEATS. Este último mostró el mejor rendimiento en términos de error y estabilidad, por lo que fue seleccionado como modelo definitivo. Su arquitectura permitió trabajar con series univariadas sin necesidad de descomponer manualmente la información, reduciendo así la complejidad del proceso y facilitando la generación de predicciones por municipio y tipo de servicio.

El desarrollo del análisis también enfrentó desafíos importantes. Uno de ellos fue la magnitud del dataset, que superó los 11 millones de registros en algunos archivos. Esto exigió un manejo cuidadoso de los recursos computacionales para evitar errores por sobrecarga de memoria o tiempos excesivos de procesamiento. Asimismo, la falta de claridad semántica en varias variables dificultó su interpretación y transformación, debido a la ausencia de documentación o glosarios que explicaran con precisión su función dentro del sistema de información.

Estos retos evidencian la necesidad de contar con datos bien estructurados y acompañados de metadatos que orienten su uso en procesos analíticos. La calidad y

trazabilidad de los datos son condiciones fundamentales para implementar soluciones predictivas que respalden decisiones en salud pública. A pesar de estas limitaciones, el modelo logró generar resultados coherentes y diferenciados por territorio, lo que refuerza su utilidad como herramienta de soporte en la gestión de la demanda de servicios médicos laborales.

8. REFERENCIAS

Shaw, J. A., Sethi, S., & Sharma, R. (2022). Addressing data quality issues in electronic health records: A multi-step approach. *Journal of Biomedical Informatics*, 128, 104015. Recuperado de: <https://doi.org/10.1016/j.jbi.2022.104015>

Jiménez, W.T. (2005). Modelo de simulación prospectiva de la demanda de servicios de salud para enfermedades de alto costo: aplicación para una entidad promotora de salud colombiana. Recuperado de: <https://studylib.es/doc/6937689/modelo-de-simulaci%C3%B3n-prospectiva-de-la-demanda-de-servici...?p=2>

Ramos, M. I., Cubillas, J. J., Jurado, J. M., Lopez, W., Feito, F. R., Quero, M., & Gonzalez, J. M. (2019). Prediction of the increase in health services demand based on the analysis of reasons of calls received by a customer relationship management. *The International journal of health planning and management*, 34(2), e1215–e1222. Recuperado de: <https://doi.org/10.1002/hpm.2763>

Stack Exchange. (2022). Which preprocessing is the correct way to forecast time-series data using LSTM? Recuperado de <https://ai.stackexchange.com/questions/40752>

Oreshkin, B. N., Dudek, G., Petka, P., & Turkina, E. (2020). N-BEATS neural network for mid-term electricity load forecasting. *arXiv*. <https://arxiv.org/abs/2009.11961>

Odhiambo, S. O., Cornelious, N. O., & Waititu, H. (2024). Developing a Hybrid ARIMA-XGBOOST Model for Analysing Mobile Money Transaction Data in Kenya. *Asian Journal of Probability and Statistics*, 26(10), 108–126. <https://doi.org/10.9734/ajpas/2024/v26i10662>

Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv*. <https://arxiv.org/abs/1905.10437>

García, B.Á. (2011). Especificación y validación de modelos de demanda de asistencia sanitaria, absentismo laboral y actitudes de los desempleados: aplicación al caso español. Recuperado de: <https://dialnet.unirioja.es/servlet/tesis?codigo=2161>

Gallego, J. M. (2008). Demanda por seguro de salud y uso de servicios médicos en Colombia: diferencias entre trabajadores dependientes e independientes. *Lecturas De Economía*, 68(68), 95–120.

Recuperado de: <https://doi.org/10.17533/udea.le.n68a266>

Verdonck, T., Baesens, B., Óskarsdóttir, M. *et al.* Special issue on feature engineering editorial. *Mach Learn* **113**, 3917–3928 (2024).

Recuperado de: <https://doi.org/10.1007/s10994-021-06042-2>