

A Barrera. | 1. Dirección de modelos de gestión de riesgo y poblaciones, Servicios de Salud IPS Suramericana  
J Zuluaga. | 2. Dirección de gestión de información y analítica, Servicios de Salud IPS Suramericana.

# Analítica predictiva para la prestación de los servicios de salud.

## Objetivo

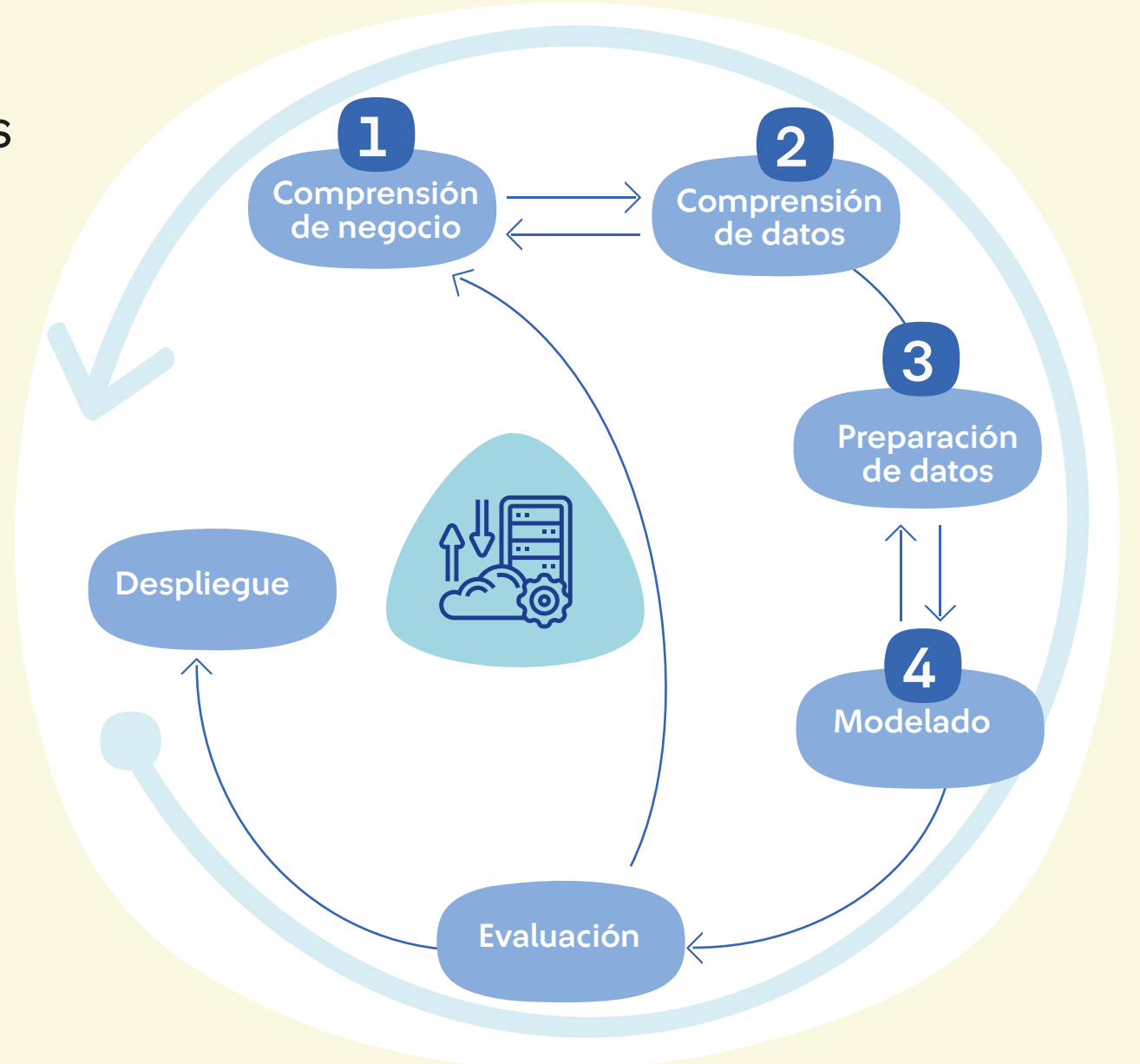
Desarrollar una estrategia de datos que maximice el valor de la información clínica y de aseguramiento en salud mediante la creación de modelos analíticos basados en aprendizaje automático, capaces de identificar patrones ocultos y apoyar la toma de decisiones clínicas y operativas para mejorar los resultados en salud y generar valor.

## Introducción

El uso de los registros médicos electrónicos ha aumentado, de acuerdo con las proyecciones realizadas por la Corporación Internacional de Datos, la tasa de crecimiento anual compuesto de los datos será del 36% hasta el 2025, y para el 2020 se crearán aproximadamente 270 GB de datos de salud por cada persona en el mundo<sup>1,2</sup>. Este crecimiento en el sector salud, ha generado un aumento en la demanda de uso y la necesidad de construir evidencia en el mundo real de la aplicabilidad de la información obtenida. La inteligencia artificial permite el procesamiento rápido, la integración, validación y análisis a gran escala de los registros médicos, y genera soluciones capaces de predecir de manera personalizada diferentes desenlaces. Por lo anterior, los modelos analíticos tienen múltiples aplicaciones, quizás las más usadas están en relación con procesos diagnósticos; sin embargo, el potencial de las herramientas basadas en modelos de inteligencia artificial ha hecho que los médicos e investigadores estén cada vez más interesados en utilizarlos para revolucionar la prestación de servicios de salud y la educación médica.

## Metodología

Para la construcción de los modelos analíticos se aplicó una técnica de minería de datos denominada CRISP-DM (de las siglas en inglés, Cross-Industry Standard Process for Data Mining) garantizando un enfoque sistemático y eficiente. Esta estructura metodológica se basa en el ciclo PHVA (Planificar - Hacer - Verificar - Actuar) y define una secuencia de 6 pasos que es flexible y puede personalizarse de acuerdo con las necesidades. (Ver gráfica 1).



Gráfica 1 | Minería de datos metodología CRISP-DM

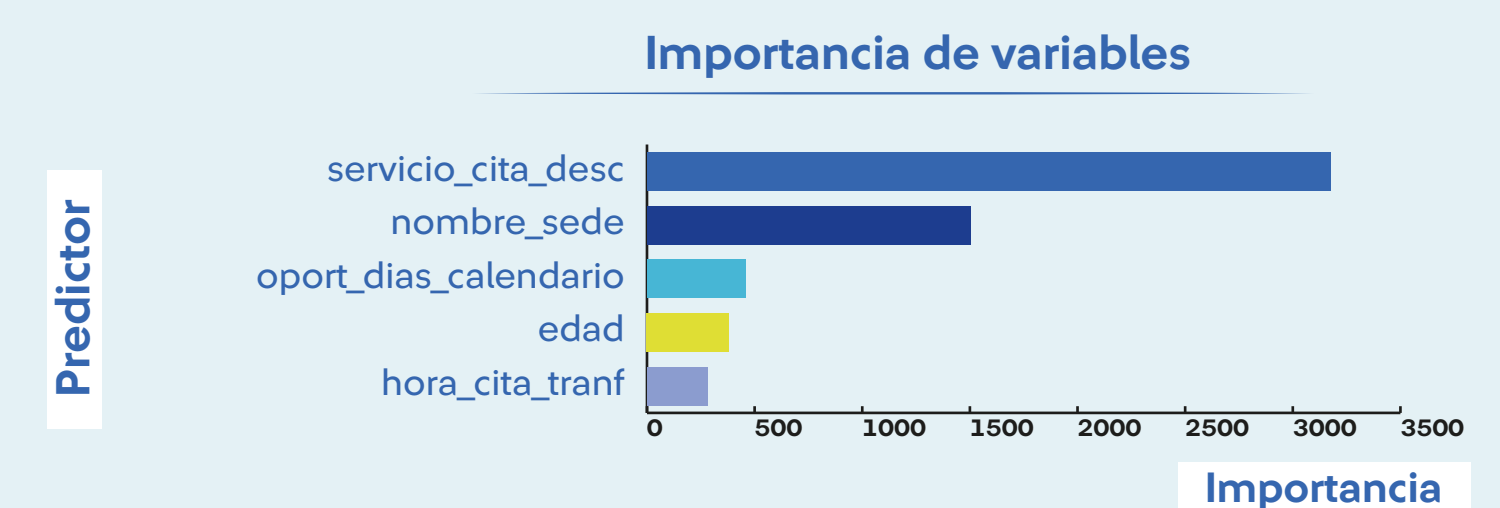
## Resultados

La metodología anteriormente descrita ha sido aplicada por la organización para la creación de diferentes modelos analíticos, entre estos tenemos:

### Modelo de inasistencias

El objetivo del modelo fue identificar citas con alta probabilidad de ser inasistidas para optimizar la gestión de agendas. Se empleó la técnica de aprendizaje supervisado LightGBM (de las siglas, Light Gradient Boosting Machine). Se utilizaron los datos de citas agendadas entre septiembre de 2022 y septiembre de 2023, así como variables sociodemográficas, antecedentes personales y patrones de utilización de los servicios (Ver gráfica 2). Se recopilaron datos de 4'986.310 citas agendadas, identificando una tasa de inasistencia del 19,07%. **El AUC de este modelo es 0.75**

La implementación de este modelo permite desplegar estrategias para reducir la tasa de inasistencia a diferentes servicios de atención en salud, optimizar el proceso de asignación y mejorar la experiencia de los usuarios.



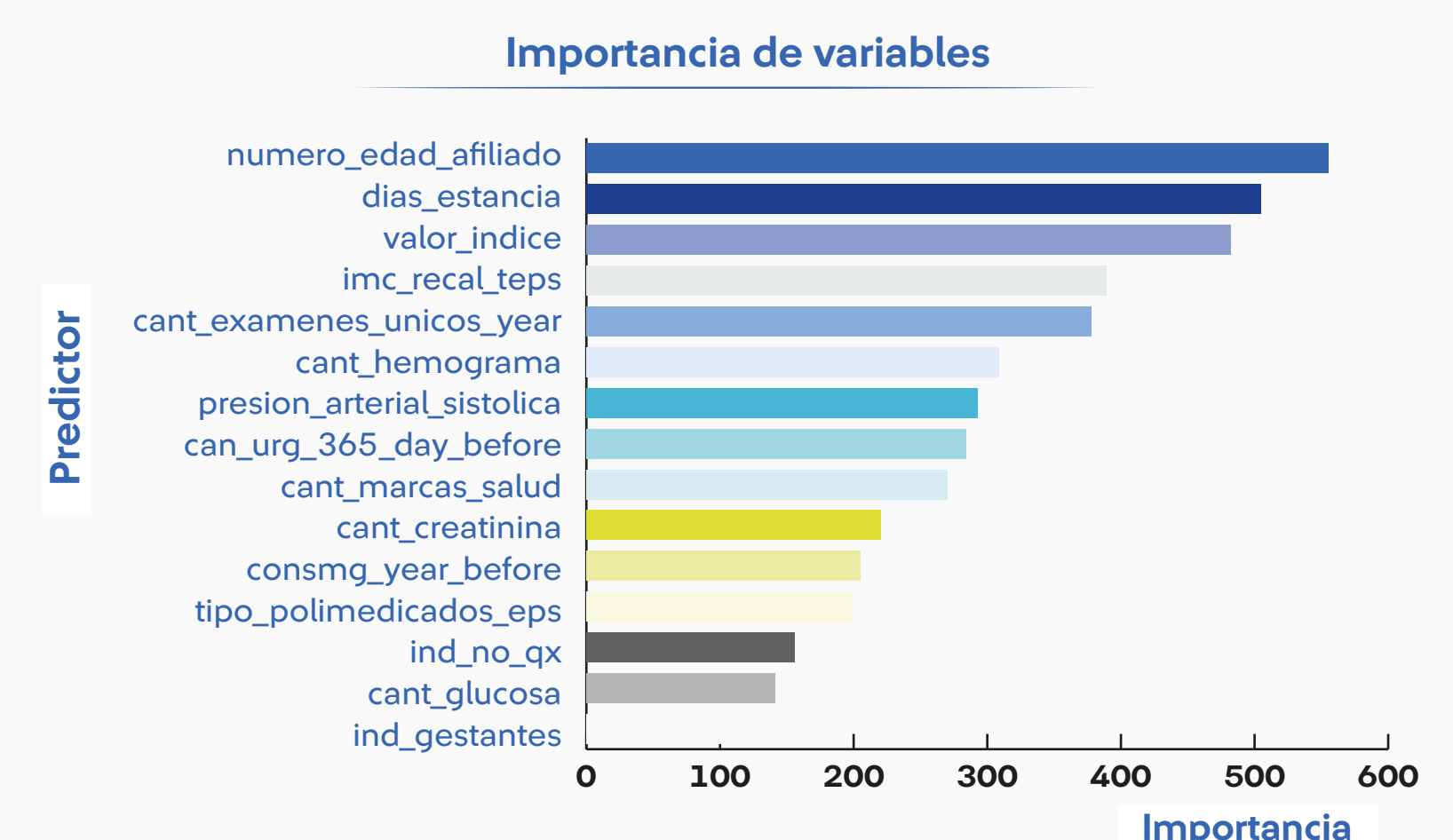
Gráfica 2 | Variables asociadas al modelo de inasistencia

### Modelo de probabilidad de re-hospitalización

El objetivo de esta solución fue diseñar un modelo preciso y confiable para la identificación de egresos con alto riesgo de reingreso hospitalario utilizando información de datos longitudinales. Se definió el uso del modelo LightGBM, para analizar los datos clínicos y sociodemográficos de los pacientes, incluyendo historial médico, patrones de utilización de servicios en salud y factores sociales relevantes (Ver gráfica 3). Se definió como reingreso un segundo evento hospitalario presentado 14 días posterior al alta. Se utilizaron datos de eventos hospitalarios ocurridos entre el 15 de diciembre de 2021 y el 31 de diciembre de 2022.

La información extraída se dividió en tres subconjuntos: entrenamiento (75%), validación (20%) y prueba (5%). Se utilizaron datos de **264.064 egresos** (40% mujeres, 60% hombres, 22% de 65 años o más) para el entrenamiento y los modelos de evaluación: **19.013 (7,2%)** fueron readmitidos dentro de los 14 días posteriores al alta por cualquier motivo. **El AUC para este modelo fue de 0.75.**

La identificación temprana de los pacientes con alto riesgo de reingreso, permitió la implementación de intervenciones preventivas para reducir la materialización de este riesgo.



Gráfica 3 | Variables asociadas al modelo de probabilidad de re-hospitalización

### Modelo de riesgo en salud en población con diabetes mellitus (DM)

El objetivo del modelo fue clasificar a los pacientes con DM por nivel de riesgo para optimizar la gestión y apalancar los objetivos de una IPS especializada en diabetes. Se implementó el modelo LightGBM.

En una ventana de tiempo de 1 año, se extrajo información de pacientes afiliados en diciembre de 2021, aplicando algunas exclusiones. El conjunto de datos se dividió en tres subconjuntos: entrenamiento (75%), validación (20%) y prueba (5%). Para evaluar la capacidad predictiva del modelo, se utilizó el **coeficiente de determinación (R2) arrojando un resultado de 0.62.**

Este modelo generó nuevos conocimientos acerca de los datos que pueden ser utilizados para identificar factores de riesgo no clásicos asociados al riesgo cardiovascular, contribuyendo al desarrollo de mejores estrategias preventivas.

Para todos los modelos se utilizó el análisis clúster de k-means, que agrupa objetos en k grupos basándose en sus características. Para nuestra organización, se distribuye la población en 5 según su nivel de riesgo.

## Conclusiones

- Nuestra experiencia nos ha permitido demostrar que el uso de la inteligencia artificial y los modelos analíticos en salud permiten mejorar la gestión del riesgo de los usuarios. A pesar del volumen y la complejidad de los datos existentes en el ecosistema de salud, es factible diseñar y desarrollar una metodología de preparación de datos orientada al dominio salud.
- Estos modelos y sus resultados son prometedores, para nuestra organización, la información obtenida permite tener una nueva perspectiva y nos ha ayudado en los procesos de toma de decisiones alrededor de la atención primaria en salud y los modelos especializados.
- La metodología presentada puede ser aplicable y útil en otras organizaciones, y puede ajustarse de acuerdo con las características de otras organizaciones.

## Bibliografía

- IDC. (Junio de 2021). The Data Dilemma and Its Impact on AI in Healthcare and Life Sciences. Obtenido de <https://blogs.idc.com/2021/06/23/the-data-dilemma-and-its-impact-on-ai-in-healthcare-and-life-sciences/#:~:text=IDC%20estimates%20that%20on%20average,establishing%20a%20data%20driven%20culture.>
- IDC. (Octubre de 2023). IDC FutureScape: Worldwide Healthcare Industry 2024 Predictions. Obtenido de <https://www.idc.com/research/viewtoc.jsp?containerId=US50105223>