



Machine Learning aplicado a la predicción de pacientes en EPS: una revisión de literatura

*Machine Learning Applied to Patient Prediction
in EPS: a Literature Review*

Johnny Urrego¹, Juan Acosta²

Tipo de Artículo: revisión de literatura.

Recibido: 08/11/2023. **Aprobado:** 10/12/2024. **Publicado:** 12/12/2024

Resumen: Esta revisión sistemática de la literatura tiene como objetivo comprender los conceptos clave, así como los estudios previos sobre *Machine Learning* (ML) y *Business Intelligence* (BI) orientados a la predicción de una variable objetivo. La metodología utilizada incluyó una exhaustiva búsqueda en las bases de datos Scopus y ScienceDirect, seleccionando aquellos estudios que cumplieran con criterios de inclusión predefinidos. Este artículo ofrece una revisión general del ML, la Inteligencia Artificial aplicada y sus principales modelos. Los hallazgos destacan tanto avances prometedores como desafíos pendientes, lo que puede servir de base para futuras investigaciones y aplicaciones en el sector salud.

Palabras clave: *Machine learning*; predictivo; salud; crecimiento; bussines intelligence.

Abstract: A systematic review of literature in which we seek to understand the terms, in addition to previous studies on *Machine Learning* (ML) and *Business Intelligence* (BI) for the prediction of an objective variable. The methodology used included an exhaustive search of the scientific literature in Scopus and ScienceDirect databases and studies were selected that met predefined inclusion criteria. This systematic literature review article provides an overview of applied ML and AI and its models. The findings highlight both promising advances and remaining challenges, which can serve as a basis for

¹ Autor correspondiente: Johnny Urrego Piedrahita Mayor título: Prof. en Gerencia de Sistemas de Información en salud. Filiación institucional: Universidad Católica Luis Amigó País: Colombia, Ciudad: Medellín. Correo electrónico: johnny.uregopi@amigo.edu.co

² Autor correspondiente: Juan José Acosta Jiménez Mayor título: Contador / Regente Farmacia. Filiación institucional: Universidad Católica Luis Amigó. País: Colombia, Ciudad: Medellín. Correo electrónico: juan.acostaim@amigo.edu.co

future research and applications in the healthcare sector.

Keywords: *Machine learning*; predictive; health; growth; *business intelligence*.

I. Introducción

Abordar la creciente necesidad de gestionar eficaz y eficientemente los recursos en el sector salud, con el fin de prever con precisión cuántos pacientes se afiliarán a un plan, puede contribuir a evitar la pérdida de recursos, optimizar la distribución del personal y brindar una atención de mayor calidad. Para esta Revisión Sistemática de la Literatura (RSL), resulta indispensable conocer el comportamiento poblacional y predecir el aumento o disminución de una población, lo que permitiría anticipar posibles posible gasto o inversiones y, en consecuencia, estimar la rentabilidad. Estos análisis se fundamentarían en modelos estadísticos basados en *machine learning*. Por ello, es necesario realizar una RSL clara, concisa y rigurosa sobre las metodologías de predicción poblacional, las posibles variables requeridas por los algunos modelos a aplicar y las formas de implementación.

Esta revisión tiene como objetivo identificar modelos predictivos que utilicen datos históricos, factores demográficos y otras variables, con el fin de proporcionar a las instituciones de salud una herramienta valiosa para la planificación y la toma de decisiones informadas. Esto permitiría mejorar, en última instancia, la calidad de la atención médica, optimizar los recursos y aumentar la satisfacción de los pacientes; además, la revisión aportará evidencia real y concreta sobre las variables que pueden incluirse o descartarse métodos estadísticos que deben implementarse y los datos necesarios para formular un modelo con resultados claros, ajustados, relevantes y creíbles.

Se realizó una búsqueda de publicaciones científicas en revistas indexadas en ScienceDirect y Scopus. Esta búsqueda, sin restricción de idioma, se enfocó en las motivaciones, metodologías de trabajo, desafíos y beneficios que han experimentado

organizaciones del sector salud al implementar la analítica de datos como estrategia de gestión organizacional.

La inteligencia artificial (IA) y el *machine learning* (ML) han dado origen a un nuevo campo en el ámbito de la atención médica y la investigación en ciencias de la salud. Estas innovadoras herramientas de análisis predictivo han comenzado a influir en diversas áreas de la atención clínica, incluida la práctica médica.

Muchas de estas herramientas y estudios basados en ML también están ganando presencia en la literatura científica, ya que la falta de conocimiento general sobre los conceptos básicos y los elementos fundamentales de la IA y el ML se ha convertido en una necesidad creciente. Esto resalta la importancia de educar al público en estos conceptos, que aún resultan relativamente desconocidos [1].

II. Justificación de la revisión

Las técnicas de aprendizaje profundo y aprendizaje automático ofrecen oportunidades inigualables para mejorar la atención sanitaria; sin embargo, existe escasez de literatura sobre aplicaciones basadas en IA implementadas con ese propósito [2].

Los datos, la información y el conocimiento han incrementado exponencialmente su impacto en la toma de decisiones, la implementación de nuevas estrategias y el crecimiento organizacional. La aplicación de herramientas de aprendizaje automático (ML / IA) a nivel empresarial puede mejorar el registro, procesamiento, seguridad y análisis de datos para generar conocimiento. Esto permite predecir y comprender el comportamiento de diversas variables y favorecer una mejora continua en los procesos.

El uso de herramientas de ML ofrece múltiples beneficios. Por ejemplo, puede ayudar a las organizaciones a identificar patrones y tendencias ocultas en grandes conjuntos de datos, lo que contribuye a la toma de decisiones más informadas y estratégicas. Además, el ML puede mejorar la eficiencia

operativa mediante la automatización de tareas repetitivas y la optimización de procesos.

La implementación de herramientas de ML también puede contribuir a la seguridad y protección de la información sensible. Mediante el uso de la IA, las organizaciones pueden detectar y prevenir posibles amenazas cibernéticas, identificar comportamientos anómalos y fortalecer sus sistemas de seguridad.

Es importante resaltar que el éxito de las herramientas de ML depende de varios factores, como la disponibilidad de datos de calidad, la infraestructura tecnológica adecuada, el compromiso gerencial y la capacitación del personal.

Existen proyectos en los que se programaban citas mediante un agente virtual *chatbot*, ofreciendo a los pacientes un nuevo canal de comunicación para la interacción del usuario. Este es un claro ejemplo de la aplicabilidad de la inteligencia artificial y el *big data* en el área de la salud.

Es fundamental garantizar la privacidad y la ética en el manejo de los datos.

Mediante esta revisión sistemática de la literatura, se busca analizar cómo se han implementado herramientas de IA y de BI en la predicción poblacional de las EPS, considerando diversas variables como las condiciones socioeconómicas, y evaluando su influencia en dichas predicciones.

III. Formulación de las preguntas de investigación

Las siguientes preguntas se plantean con el fin de conocer las herramientas y enfoques implementados a nivel de IA y BI en los comportamientos poblacionales.

1. ¿Qué variables pueden utilizarse para la predicción de comportamientos poblacionales de usuarios de servicios de salud?

2. ¿Cómo se ha aplicado el ML en la predicción de comportamientos poblacionales de usuarios de servicios de salud?

3. ¿Cómo se ha aplicado BI para mejorar la gestión de las EPS?

IV. Definiciones y conceptos básicos

ARIMA: técnica de minería de datos que ofrece resultados confiables, robustos y de calidad, ayudando a validar y sustentar investigaciones mediante mecanismos eficientes para la extracción, análisis y visualización de grandes volúmenes de datos. Además, presenta la gran ventaja de ser aplicable en múltiples ramas del conocimiento, como economía, demografía, física, matemáticas, pesca, entre otras [3].

BI: Business Intelligence: es una disciplina y conjunto de herramientas que integran datos operativos con análisis avanzado para proporcionar información compleja y competitiva a planificadores y tomadores de decisiones. Las herramientas de inteligencia empresarial (BI) ayudan a las organizaciones a generar rápidamente conocimientos que guían a los gerentes hacia la eficiencia operativa, los conducen a nuevas oportunidades y los diferencian de la competencia [4].

Big Data: término amplio y genérico que abarca varias características distintas y no exclusivas. El origen del concepto se atribuye comúnmente a un breve informe técnico del grupo META, una consultora estadounidense, y evolucionó hacia el modelo de las "3V: volumen, variedad y velocidad. Posteriormente, se han propuesto otras "V", como veracidad. Todas estas "V" se refieren a datos. Los desafíos que enfrenta la sociedad de la información están relacionados con el procesamiento en tiempo real de grandes volúmenes de datos diversos y heterogéneos (Velocity) [5].

Crecimiento Poblacional: se refiere al cambio en el tamaño de una población durante un período de tiempo específico. Se puede medir como la

variación en el número de individuos por unidad de tiempo [6].

EPS: según el artículo 177 de la Ley 100 de 1993, se definen como las entidades promotoras de salud responsables de los afiliados y el recaudo de sus cotizaciones, por delegación del Fondo de Solidaridad y Garantía [7].

IA: posiblemente la mayor revolución tecnológica del siglo XXI. Sus orígenes se sitúan a finales de los años 50, cuando la ciencia y la ingeniería se unieron para crear máquinas inteligentes, especialmente programas informáticos inteligentes. Se define como “la capacidad de dar buenas respuestas desde el punto de vista de los hechos” [8].

Machine Learning: el concepto de *Machine Learning* (ML) o «aprendizaje máquina», data de mediados del siglo XX y fue definido en un artículo de Samuel en 1959 como una rama de la IA que utiliza técnicas estadísticas y algoritmos computacionales para dotar a los ordenadores de la capacidad de «aprender», es decir, mejorar su desempeño en una tarea específica tras procesar grandes volúmenes de datos, sin instrucciones explícitas externas (y, por tanto, potencialmente sesgadas) proporcionadas por el programador [9].

V. Proceso de búsqueda de documentos

En el proceso de búsqueda de documentos relacionados con la RSL, se analizó la producción científica en las bases de datos bibliográficas Scopus y ScienceDirect. La búsqueda no tuvo restricciones de idioma ni de ubicación, y se seleccionaron artículos publicados en los últimos cuatro años que contuvieran los términos de búsqueda en inglés. De este modo, se identificó el idioma más utilizado entre las publicaciones.

Para esta actividad se utilizaron las siguientes palabras clave, variables, *prediction*, *predictive*, *machine learning*, *learning system*, *learning algorithms*, *health*, *population*, *growth*, *Big Data*, combinadas por “and”, para ampliar el alcance de la búsqueda.

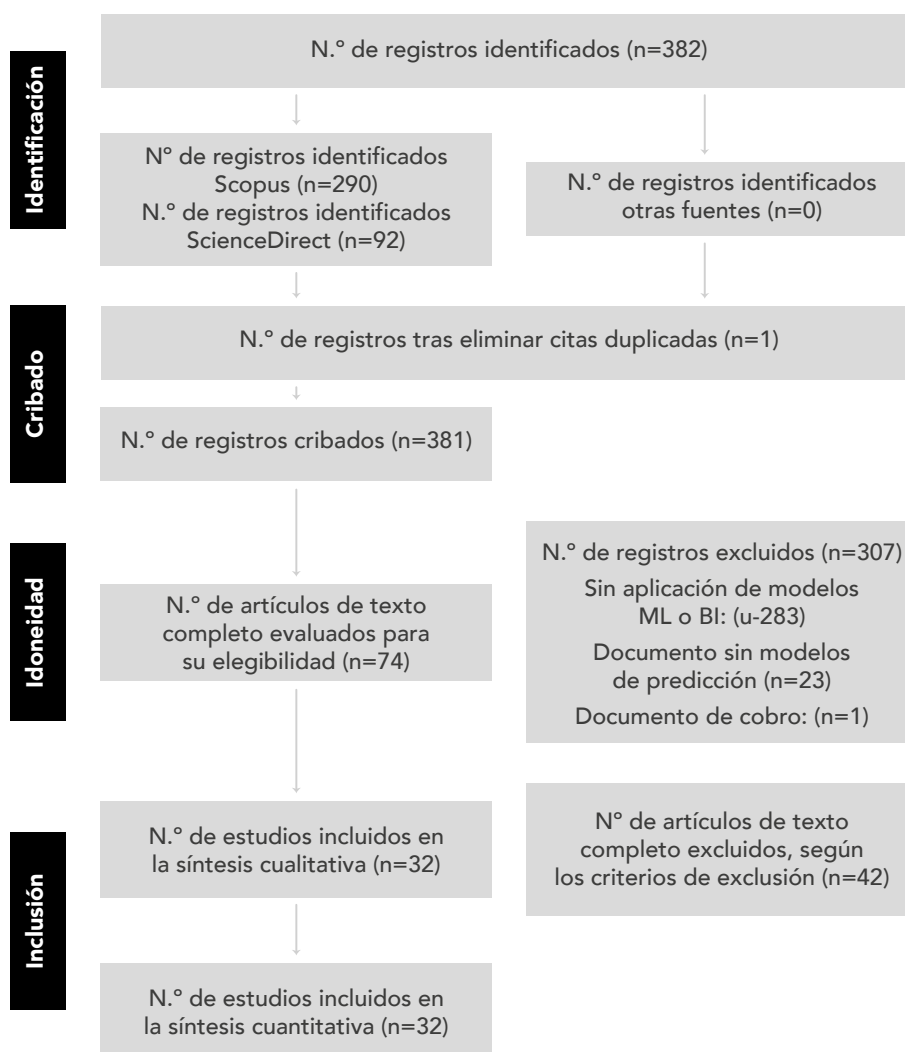
Posterior a la búsqueda inicial, se aplicaron filtros como año de publicación, tipo de documento, temáticas, acceso abierto (*open source*) y, de gran importancia, el filtro de “keyword” (palabra clave), mediante el cual el buscador ofrece las opciones óptimas para afinar la búsqueda relacionada.

En la utilización de la “keyword” se identificaron términos adicionales para la búsqueda, tales como *population statistics*, *artificial intelligence*, *predictive value*, *predictive analytics*, *population research*, *models*, *statistical*, *predictive models*, *machine learning models*, *prediction modelling* y *prediction model*.

Esto arrojó 92 resultados en ScienceDirect y 290 resultados en Scopus con los filtros aplicados.

De los artículos encontrados mediante las palabras clave y los parámetros de búsqueda, se seleccionaron aquellos relacionados directamente con el tema principal y cuya redacción fuera específica, clara y coherente. Para ello, se llevó a cabo una lectura de los resúmenes con el fin de identificar los principales resultados y conclusiones que contribuyen a la resolución del objetivo planteado.

Se excluyeron los documentos en los que no se explicitaban los modelos de predicción, así como aquellos artículos de carácter exclusivamente asistencial que no presentaban evidencia del uso de herramienta de BI o ML. También se destacaron aquellos trabajos en los que no se aplicó ninguna técnica de aprendizaje automático.

Figura 1. Flujograma Prisma.

Nota: fuente, consulta de artículos.

VI. Discusión

Pregunta 1

¿Qué variables se han planteado o se han considerado utilizar para la predicción de comportamientos poblacionales de usuarios de servicios de salud?

El estudio de van Hoffen [10], centrado en el desarrollo de modelos de predicción de ausencias laborales por enfermedad mental en la población activa general, investigó si las variables obtenidas en encuestas de salud ocupacional pueden emplearse, y de qué manera para identificar a los trabajadores en riesgo de incapacidad prolongada por trastornos mentales. A partir de una consulta médica, se registra un código diagnóstico basado en la décima Clasificación Internacional de Enfer-

medades (CIE-10), y se realiza un seguimiento durante un año como variable de resultado.

Los autores, Verma et al. [11] argumentan que actualmente no existe un marco claro e integral para la adopción de herramientas basadas en aprendizaje automático de la salud. Por ello, proponen un modelo de adopción desarrollado en tres fases: exploración, diseño de soluciones e implementación y evaluación. Asimismo, sugieren el uso de marcos metodológicos como CRISP DM 7 para el desarrollo y validación de algoritmos, destacando como una ventaja principal de este enfoque su capacidad para incorporar la naturaleza iterativa del proceso de ciencia de datos.

Con el objetivo de definir la variable de movilidad geográfica entre los participantes en la búsqueda de casos de tuberculosis en centros y comunidades de Kampala, Uganda, Robsky et al. [12] evaluaron las asociaciones entre la movilidad, el acceso a la atención y la enfermedad de tuberculosis, incluyendo desplazamientos por motivos laborales, educativos o personales. Estas variables se consideraron relevantes en la dinámica de transmisión de la enfermedad. El estudio encontró una asociación entre la movilidad y un menor riesgo de desarrollar tuberculosis, lo cual podría explicarse, en parte, por un estatus socioeconómico más alto entre las personas móviles de esta población.

Las variables utilizadas para alimentar el modelo ML incluyeron cuatro parámetros climáticos: (temperatura, humedad, presión atmosférica y velocidad del viento) y dos indicadores de contaminación del aire (PM2,5 y PM10). Estos se emplearon como datos de entrada para predecir tanto la infección como la recuperación. Dichas variables pueden considerarse factores adicionales que contribuyen a enfocar y entrenar el modelo, permitiendo una predicción más precisa [13].

Es importante conocer las estadísticas nacionales de los indicadores demográficos y de salud con el objetivo de desarrollar modelos de predicción de la multimorbilidad. Entre las técnicas aplicadas, se destaca el uso de redes neuronales con retro propagación. Para ellos, se emplean como predictores

las variables identificadas mediante modelos de regresión logística, tales como el género, el hábito de fumar, la edad y el consumo de frutas, las cuales desempeñan un papel importante en la capacidad predictiva del modelo [14].

Los sistemas ciber físicos (CPS) integran capacidades de computación y comunicación para regular procesos físicos y facilitar la interacción entre los entornos digital y físico en tareas de control y monitoreo. En el ámbito de la salud, una variante específica conocida como Health CPS se emplea como sistema de monitoreo de condiciones médicas. Este tipo de sistema permite la captura, el procesamiento y el análisis dinámico de datos provenientes de sensores de salud mediante procesos biofísicos integrados, habilitados por tecnologías del Internet de las cosas (IoT) [15].

Los investigadores Mbunge y Batani [2] utilizaron múltiples sistemas y aplicaciones basadas en inteligencia artificial para mejorar la atención médica en el África subsahariana, con énfasis en algoritmos de aprendizaje profundo y aprendizaje automático. Estas técnicas de inteligencia artificial se han utilizado principalmente en tareas como la predicción de brotes epidémicos, la vigilancia de enfermedades, la evaluación y diagnóstico de riesgos, la provisión de atención personalizada en tiempo real, el monitoreo remoto de pacientes y la atención preventiva.

En el artículo publicado por Alberto et al. [16], sobre el impacto de los conjuntos de datos comerciales de salud en la investigación médica y los algoritmos de atención médica, se identifican variables clave para la aplicación de inteligencia artificial en estudios relacionados con la salud, tales como atenciones médicas, clínicas y uso de fármacos. Además, los autores destacan la relevancia de diversos repositorios de datos, tanto abiertos como privados, para la investigación en esta área; por ejemplo, la base de datos del laboratorio de Fisiología Computacional del Instituto Tecnológico de Massachusetts, así como el repositorio de códigos de MIMIC, que ha sido utilizado en más de 4.000 estudios basados en registros médicos electrónicos de un hospital en Estados Unidos.

Hemos ingresado en una nueva era digital en el campo de la medicina, caracterizada por el potencial de aprovechar grandes volúmenes de información médica para acelerar los descubrimientos científicos y mejorar la calidad de la atención médica. Este avance permite, además, la prestación de servicios médicos personalizados y respalda la formulación de políticas fundamentadas en evidencia. En este contexto, marcado por la convergencia de diversos factores, los datos relacionados con la atención médica, incluidos los datos demográficos de los pacientes, los resultados de pruebas clínicas, hallazgos de laboratorio y datos genómicos, se han convertido en un activo de gran valor tanto para entidades públicas como privadas. La creciente demanda de grandes volúmenes de información médica, tanto la academia como la industria, ha impulsado una dinámica tendencia hacia su comercialización, dando lugar a una industria multimillonaria centrada en la recopilación, el análisis y la comercialización de datos [17].

El uso del análisis predictivo, respaldado por marcos basados en Internet de las cosas (IoT), ha representado un avance significativo en el desarrollo de los sistemas de salud inteligentes, al permitir la anticipación de tendencias de mortalidad y, con ello, la implementación de intervenciones oportunas. Esta metodología también facilita la identificación de relaciones ocultas entre variables de salud y factores críticos asociados a la mortalidad infantil. Para este propósito, se emplearon conjuntos de datos provenientes de la Encuesta Demográfica de Salud de Etiopía y la Encuesta Demográfica de Salud de Pakistán. Dichos datos reales fueron analizados mediante clasificadores de aprendizaje automático, tales como *Naïve Bayes*, árboles de decisión, inducción de reglas, bosque aleatorios y perceptrones multicapa, con el fin de realizar tareas de predicción [18].

Pregunta 2

¿Cómo se ha aplicado el ML en la predicción de comportamientos poblacionales de usuarios de servicios de salud?

Entre los algoritmos más utilizados para la construcción de modelos de aprendizaje automático en el ámbito de la salud se encuentran *Random Forest*, *Support Vector Machines* y las Redes Neuronales. Asimismo, técnicas de *Deep Learning* o aprendizaje profundo han sido aplicadas en desarrollos más recientes. Este enfoque ha permitido abordar tareas complejas que anteriormente requerían intervención humana, logrando resultados significativos en áreas como la predicción de enfermedades, el monitoreo de pacientes y la personalización de tratamientos. En este contexto, el aprendizaje automático brinda nuevas oportunidades para mejorar la eficiencia en la prestación de servicios de salud, al tiempo que respalda la toma de decisiones clínicas mediante la integración de grandes volúmenes de datos y la identificación de patrones relevantes [19].

Los sistemas de salud basados en IA han demostrado mejoras significativas en la predicción, el diagnóstico y las estrategias terapéuticas, generando beneficios tanto para los pacientes como para los profesionales de la salud [8]. La aplicación de la IA en la atención sanitaria puede aportar valor en al menos cuatro dimensiones: (I) estimar la probabilidad de éxito de un tratamiento y anticipar la aparición de enfermedades antes de su manifestación clínica; (II) prevenir o gestionar posibles complicaciones; (III) apoyar activamente la atención al paciente durante el proceso diagnóstico y/o tratamiento; y (IV) identificar con mayor precisión la patología subyacente y sugerir las opciones terapéuticas más adecuadas [20].

Los servicios de salud móviles impulsados por IA pueden generar avances significativos en el monitoreo y la evaluación de los pacientes, al permitir procesos más automáticos, objetivos y continuos. Además, las metodologías recientes de IA explicable (XAI) contribuyen a que las decisiones generadas por los modelos sean interpretables y confiables para los usuarios finales. Esto fortalece la

confianza en los modelos basados en IA, y en particular en el aprendizaje automático (ML), al hacer que las decisiones derivadas sean percibidas como más precisas y fundamentadas [21].

En el estudio de Mirzaeian et al. [22]. Se utilizó inicialmente la regresión logística binaria (BLR) para identificar los principales factores que influyen en el modelo. Posteriormente, se aplicaron ocho algoritmos de aprendizaje automático para realizar predicciones, entre ellos: refuerzo adaptativo (*Ada-Boost*), agregación de arranque (*Bagging*), refuerzo de gradiente extremo (*XG-Boost*), bosque aleatorio (RF), J-48, perceptrón multicapa (MLP), Naïve Bayes (NB), y máquina de vectores de soporte (SVM). El rendimiento de cada modelo se evaluó mediante métricas derivadas de la matriz de confusión, con el objetivo de determinar cuál ofrecía la mejor capacidad predictiva. Aunque algunos de estos métodos y algoritmos pueden no ser ampliamente conocidos, se emplearon eficazmente para la predicción de la esperanza de vida.

Los datos administrativos de salud longitudinales a nivel poblacional facilitan el desarrollo de un sistema sanitario de aprendizaje. A partir de estos, se han desarrollado modelos de aprendizaje automático orientados a predecir el riesgo de mortalidad en pacientes que ingresan a hospitales. Esto permite inferir que la disponibilidad y el análisis adecuado de la información son fundamentales para una toma de decisiones acertada, la focalización eficiente de inversiones y una gestión adecuada, tanto en el ámbito asistencial como administrativo, por parte de clínicas, hospitales y aseguradoras de salud [23].

Aunque en más del 80% de los casos los modelos de ML en salud pueden predecir cómo se propagará una enfermedad, aún se buscan estrategias para mejorar la precisión en la predicción del comportamiento poblacional. En el experimento presentado en el documento, el algoritmo Random Forest predijo correctamente los casos positivos de malaria en un 60% de las veces y los casos positivos de fiebre tifoidea en un 77%. Este resultado destaca a Random Forest como otro modelo potencialmente útil para tareas de predicción en el ámbito sanitario [24].

Con datos de características sociodemográficas, conductuales (como el uso de drogas inyectables y el comportamiento sexual) y biológicas, el estudio de Nisa [25] realizó un análisis orientado a predecir pacientes con alto riesgo de contraer VIH. Además, se reporta el uso de técnicas de minería de datos y aprendizaje automático, lo que permite identificar qué tipo de variables pueden considerarse relevantes para el ajuste de los modelos predictivos.

En la metodología del aprendizaje automático en informática de la salud (SEMLHI) incluye cuatro módulos: *software*, aprendizaje automático, algoritmos de aprendizaje automático y datos de informática sanitaria. Estos módulos organizan las tareas dentro del marco SEMLHI, lo que permite a investigadores y desarrolladores analizar el *software* de informática en salud desde una perspectiva de ingeniería. Asimismo, proporciona una hoja de ruta innovadora para el diseño de aplicaciones de salud, incorporando funciones de sistema e implementaciones de *software* más eficaces [1].

Pregunta 3

¿Cómo se ha aplicado BI para mejorar la gestión de las EPS?

El trabajo de investigación de Kcomt-Ponce et al. [26] desarrolló una aplicación móvil que integra inteligencia artificial, *machine learning* e inteligencia de negocios para mejorar el proceso de asignación de citas médicas. Además, la aplicación incluye funcionalidades como la cancelación de citas, la revisión de la historia clínica y la confirmación de asistencia al centro de salud correspondiente, contribuyendo así a una gestión más eficiente y centrada en el usuario.

Un ejemplo destacado de la aplicación de *Big Data* e inteligencia de negocios en el área de la salud se encuentra en la investigación de Mohapatra et al. [27], donde se emplean teléfonos inteligentes y microsensors integrados en dispositivos portátiles para el monitoreo remoto de signos vitales. En este estudio, se propone un marco para la detección oportuna de situaciones de emergencia,

basado en la sensibilidad del modelo, que permite la transmisión eficiente de los datos captados por los sensores, así como su recopilación y análisis en tiempo real.

Una herramienta de gestión de la salud poblacional, basada en algoritmos inteligentes de última generación y alimentada por datos administrativos y socioeconómicos, ha pronosticado que, para el año 2060, una cuarta parte de la población estadounidense tendrá más de 65 años. Esta tendencia no será exclusiva de Estados Unidos, ya que se proyecta un crecimiento similar en otros países. Además, se estima que la población de 85 años o más casi se duplicará en los próximos 25 años. En este contexto, los procesos automatizados de *Big Data* para definir cohortes de pacientes y para estratificar grupos por nivel de riesgo, buscan mejorar los resultados clínicos y reducir los costos, contribuyendo así a una gestión más eficiente de las organizaciones de salud [28].

Un estudio realizado por el grupo Keralty Salud Colombia, enfocado en la previsión de costos médicos en empresas de salud mediante el uso de modelos basados en analítica avanzada, desarrolló una estrategia efectiva para la predicción de costos. Esta estrategia contribuyó a la optimización de recursos y a una toma de decisiones más informada, mediante el empleo de técnicas como el modelo de memoria a largo plazo (LSTM) y la agrupación de pacientes utilizando el algoritmo de *k-means clustering* [29].

La búsqueda de una utilización eficiente de los activos ha sido una constante, especialmente en industrias con procesos de mecanización en evolución. En este contexto, la gestión del mantenimiento adquiere relevancia, ya que es responsable de garantizar la disponibilidad operativa de dichos activos. El mantenimiento predictivo (PDM) se ha consolidado como una de las principales estrategias en este ámbito, al permitir la detección temprana de fallas, evitando paradas no programadas y costos innecesarios. Con el avance de las tecnologías, el PDM ha evolucionado hacia la Gestión de Pronóstico y Salud (PHM), que ofrece herramientas para el reconocimiento de patrones,

la identificación de anomalías y la estimación de la vida útil restante (RUL) del equipo [30].

El autor Montes [31] realizó un estudio sobre la estimación de la duración del servicio mediante algoritmos de aprendizaje automático y redes neuronales, enfocado en pacientes que reciben atención médica domiciliaria. La investigación se realizó en 14 hospitales que prestan este tipo de servicios en la ciudad de Diyarbakır, Turquía. Los datos obtenidos fueron sometidos a un preprocesamiento y análisis estadístico descriptivo. Se aplicaron varios modelos de predicción, como árbol de decisión, bosque aleatorio y red neuronal de perceptrón multicapa, con el fin de identificar el modelo más adecuado. Los resultados mostraron que la duración del servicio de atención médica domiciliaria variaba según la edad y el sexo de los pacientes, quienes en su mayoría pertenecían a grupos con enfermedades que requerían tratamiento de Fisioterapia y Rehabilitación. Se determinó que la duración del servicio puede predecirse con alta precisión utilizando algoritmos de aprendizaje automático: modelo multicapa: 90,4%, modelo de árbol de decisión: 86,4% y modelo de bosque aleatorio: 88,5% [31].

Con el objetivo de predecir inasistencias a citas médicas en un hospital pediátrico de Chile mediante técnicas de aprendizaje automático, un grupo de investigadores desarrolló una plataforma computacional que implementa modelos predictivos a través de una aplicación web. El *front-end* y el *back-end* fueron diseñados en Python usando el *framework web* Django. La entrada del sistema consiste en una hoja de cálculo que contiene las características de la cita, tales como la identificación del paciente, información personal, especialidad médica, fecha y hora. Estos datos se procesan para generar las variables específicas para cada especialidad. Posteriormente, se aplican modelos predictivos, seleccionando el de mejor desempeño, para asignar etiquetas a todas las citas. Las citas se ordenan en forma descendente según la probabilidad estimada de inasistencia, junto con la información de contacto del paciente. Esto permite al hospital contactar prioritariamente a los pacien-

tes con mayor riesgo de no presentarse, con el fin de confirmar la cita [32].

Los sensores portátiles han demostrado ser prometedores como método no intrusivo para la recopilación de biomarcadores que pueden correlacionarse con niveles elevados de estrés. Los factores estresantes desencadenan una variedad de respuestas biológicas que pueden medirse a través de estos biomarcadores, entendidos como cambios en las condiciones de salud asociados a la presencia de dichos factores. Esta información puede contribuir a una mejor prevención de enfermedades y, en consecuencia, a una reducción de los costos para las instituciones de salud. El seguimiento y monitoreo de la salud mediante dispositivos portátiles está ganando popularidad; sin embargo, la generalización de los modelos de aprendizaje automático existentes aún requiere más investigación. Se espera que los avances continúen a medida que se disponga de conjuntos de datos más amplios y actualizados [33].

El trabajo de Parro et al. [34] presenta una propuesta práctica para estimar la utilización del sistema de salud en casos de COVID-19. La metodología desarrollada se basa en el modelo dinámico SIRD (Susceptibles, Infectados, Eliminados y Muertos), el cual fue modificado para centrarse en la dinámica del sistema sanitario, en lugar de modelar todos los casos de la enfermedad. El modelo se calibró utilizando datos disponibles para cada estado brasileño y se actualizaba diariamente. Se definió y empleó una figura de mérito que evalúa la calidad del ajuste del modelo a los datos, lo que permitió optimizar los parámetros libres. Se estimaron así los parámetros de un modelo epidemiológico para todo Brasil, mediante una combinación lineal de los modelos correspondientes a los 26 estados. El modelo fue validado y mostró una alta capacidad de ajuste en la mayoría de los casos.

VII. Conclusiones

En el ámbito de la salud, los beneficios de las soluciones basadas en inteligencia artificial son tan significativos que un número creciente de ase-

guradoras y prestadores de servicios se muestra interesado en invertir en estas tecnologías. Esto evidencia no sólo una dimensión comercial emergente, sino también el creciente valor e importancia dentro del sector sanitario.

Dado que existen diversos proyectos de desarrollo en el campo de la IA capaces de generar impactos significativos y agregar valor a las EPS, el ML representa un recurso de alto potencial para el sector salud. Estos proyectos contribuyen a mejorar la eficiencia y precisión en los diagnósticos médicos, optimizar los procesos de atención al paciente y facilitar tanto la toma de decisiones clínicas como administrativas. La implementación de soluciones basadas en ML permite obtener resultados más rápidos y precisos, lo que potencialmente se traduce en una mejor calidad de atención y una reducción de costos. Asimismo, estos desarrollos permiten identificar patrones y tendencias en grandes volúmenes de datos asistenciales, lo cual favorece la investigación, el diseño de nuevos programas y/o tratamientos, y una toma de decisiones más informada y precisa.

Hoy en día, la combinación de IA (Inteligencia Artificial), *Big Data*, ML (aprendizaje automático) y BI (*Business Intelligence*) resulta fundamental para optimizar los recursos y maximizar resultados. Estas tecnologías actúan de forma complementaria para recopilar, procesar y analizar grandes volúmenes de datos, lo que permite generar información valiosa y respaldar una toma de decisiones más precisa. La IA y el ML permiten identificar patrones y tendencias ocultas en los datos, mejorando la eficiencia y eficacia de los procesos. Por su parte, *Big data* y BI proporcionan las herramientas necesarias para almacenar, gestionar y visualizar los datos de manera ágil y efectiva. Al integrar estas tecnologías, las organizaciones pueden obtener una ventaja competitiva significativa, optimizando sus recursos y elevando el rendimiento de sus procesos operativos y estratégicos.

La continuidad del desarrollo es un campo versátil y en constante evolución, cuyos beneficios apenas comienzan a explorarse en profundidad. A medida que la tecnología avanza, surgen nuevas

oportunidades que permiten la mejora continua de productos y servicios. La adopción de metodologías ágiles y enfoques de desarrollo iterativo facilita a las organizaciones adaptarse con rapidez a los cambios y ofrecer soluciones más eficientes. Aunque aún queda mucho por descubrir en este ámbito, es evidente que los futuros avances podrían tener un impacto significativo en diferentes sectores y áreas de negocio. Profundizar en los beneficios de este enfoque puede abrir nuevas oportunidades para optimizar procesos, mejorar la experiencia del usuario y aumentar la eficiencia, eficacia y efectividad organizacional.

VIII. Referencias

- [1] M. Moreb, T. A. Mohammed, and O. Bayat, "A novel software engineering approach toward using machine learning for improving the efficiency of health systems", *IEEE Access*, vol. 8, pp. 23169–23178, 2020, doi: 10.1109/ACCESS.2020.2970178.
- [2] E. Mbunge and J. Batani, "Application of deep learning and machine learning models to improve healthcare in sub-Saharan Africa: Emerging opportunities, trends and implications", *Telematics and Informatics Reports*, vol. 11, Sep. 2023, doi: 10.1016/j.teler.2023.100097.
- [3] S. Zhang, L. Tao, and T. Zheng, "Industry-Academic-Education Integration: Innovation and Practice of Training Mode of Applied Statistics Talents in Big Data Era —Taking Zhejiang University of Science and Technology as an Example", *Open J. Soc. Sci.*, vol. 09, n° 10, pp. 507–512, 2021, doi: 10.4236/JSS.2021.910037.
- [4] M. A. Tripathi, K. Madhavi, V. S. P. Kandi, V. K. Nassa, B. Mallik, and M. K. Chakravarthi, "Machine learning models for evaluating the benefits of business intelligence systems", *Journal of High Technology Management Research*, vol. 34, n° 2, 2023, doi: 10.1016/j.hitech.2023.100470.
- [5] T. Lefèvre, "Big data in forensic science and medicine", *Journal of Forensic and Legal Medicine*, vol. 57. Churchill Livingstone, pp. 1–6, 2018. doi: 10.1016/j.jflm.2017.08.001.
- [6] "Glosario de Conceptos". Accessed: Oct. 02, 2023. [Online]. Available: <https://www.ine.es/DEFIne/es/concepto.htm?c=5094&tf=&op=>
- [7] Ministerio de Salud, "Ley 100 1993".
- [8] R. B. Ruiz y J. D. Velásquez, "Artificial intelligence at the service of the health of the future", *Revista Médica Clínica Las Condes*, vol. 34, n° 1, pp. 84–91, 2023, doi: 10.1016/j.rmcl.2022.12.001.
- [9] A. Núñez Reiz, M. A. Armengol de la Hoz, and M. Sánchez García, "Big Data Analysis and Machine Learning in Intensive Care Units", *Medicina Intensiva*, vol. 43, n° 7, pp. 416–426, 2019. doi: 10.1016/j.medin.2018.10.007.
- [10] M. F. A. van Hoffen, G. Norder, J. W. R. Twisk, and C. A. M. Roelen, "Development of Prediction Models for Sickness Absence Due to Mental Disorders in the General Working Population", *J. Occup. Rehabil.*, vol. 30, no. 3, pp. 308–317, 2020, doi: 10.1007/s10926-019-09852-3.
- [11] A. A. Verma et al., "Implementing machine learning in medicine," *CMAJ*, vol. 193, n° 34, pp. E1351–E 1357, 2021, doi: 10.1503/cmaj.202434.
- [12] K. O. Robsky et al., "Characterization of geographic mobility among participants in facility- And community-based tuberculosis case finding in urban Uganda", *PLoS One*, vol. 16, n° 5, 2021, doi: 10.1371/journal.pone.0251806.
- [13] L. K. Shrivastav and R. Kumar, "Empirical Analysis of Impact of Weather and Air Pollution Parameters on COVID-19 Spread and Control in India Using Machine Learning

Algorithm", *Wirel Pers Commun*, vol. 130, n° 3, pp. 1963–1991, 2023, doi: 10.1007/s11277-023-10367-7.

- [14] F. M. Albagmi et al., "Predicting Multimorbidity Using Saudi Health Indicators (Sharik) Nationwide Data: Statistical and Machine Learning Approach", *Healthcare (Switzerland)*, vol. 11, n° 15, 2023, doi: 10.3390/healthcare11152176.
- [15] R. Ferdousi, M. A. Hossain, and A. El Saddik, "Early-Stage Risk Prediction of Non-Communicable Disease Using Machine Learning in Health CPS", *IEEE Access*, vol. 9, pp. 96823–96837, 2021, doi: 10.1109/ACCESS.2021.3094063.
- [16] I. R. I. Alberto et al., "The impact of commercial health datasets on medical research and health-care algorithms", *Lancet Digit Health*, vol. 5, n° 5, pp. e288–e294, 2023, doi: 10.1016/S2589-7500(23)00025-0.
- [17] K. M. Kim, D. S. Evans, J. Jacobson, X. Jiang, W. Browner, and S. R. Cummings, "Rapid prediction of in-hospital mortality among adults with COVID-19 disease", *PLoS One*, vol. 17, n° 7, 2022, doi: 10.1371/journal.pone.0269813.
- [18] M. Islam, M. Usman, A. Mahmood, A. A. Abbasi, and O. Y. Song, "Predictive analytics framework for accurate estimation of child mortality rates for Internet of Things enabled smart healthcare systems", *Int J Distrib. Sens Netw*, vol. 16, n° 5, 2020, doi: 10.1177/1550147720928897.
- [19] C. Aracena, F. Villena, F. Arias, and J. Dunstan, "Applications of machine learning in healthcare", *Revista Médica Clínica Las Condes*, vol. 33, n° 6, pp. 568–575, 2022, doi: 10.1016/j.rmcl.2022.10.001.
- [20] B. L. Jimma, "Artificial intelligence in healthcare: A bibliometric analysis", *Telematics and Informatics Reports*, vol. 9, p. 100041, 2023, doi: 10.1016/j.teler.2023.100041.
- [21] F. Di Martino, F. Delmastro, and C. Dolciotti, "Explainable AI for malnutrition risk prediction from m-Health and clinical data", *Smart Health*, p. 100429, 2023, doi: 10.1016/j.smhl.2023.100429.
- [22] R. Mirzaeian, R. Nopour, Z. Asghari Varzaneh, M. Shafiee, M. Shanbehzadeh, and H. Kazemi-Arpanahi, "Which are best for successful aging prediction? Bagging, boosting, or simple machine learning algorithms?", *Biomed Eng Online*, vol. 22, n° 1, 2023, doi: 10.1186/s12938-023-01140-9.
- [23] [23] W. Sun et al., "Towards artificial intelligence-based learning health system for population-level mortality prediction using electrocardiograms", *NPJ Digit Med*, vol. 6, n° 1, 2023, doi: 10.1038/s41746-023-00765-3.
- [24] M. Hussain et al., "Machine learning based efficient prediction of positive cases of waterborne diseases", *BMC Med. Inform. Decis Mak*, vol. 23, n° 1, Dec. 2023, doi: 10.1186/s12911-022-02092-1.
- [25] S. U. Nisa, A. Mahmood, F. S. Ujager, and M. Malik, "HIV/AIDS predictive model using random forest based on socio-demographic, biological and behavioral data", *Egyptian Informatics Journal*, vol. 24, n° 1, pp. 107–115, 2023, doi: 10.1016/j.eij.2022.12.005.
- [26] E. J. Kcomt-Ponce, E. L. Huamaní, and A. Delgado, "Implementation of Machine Learning in Health Management to Improve the Process of Medical Appointments In Perú", *International Journal of Emerging Technology and Advanced Engineering*, vol. 12, n° 2, pp. 74–85, 2022, doi: 10.46338/ijetae0222_09.

- [27] S. Mohapatra, A. Sahoo, S. Mohanty, and D. Singh, "IoT Enabled Ubiquitous Healthcare System using Predictive Analytics", *Procedia Computer Science*, Elsevier B.V., pp. 1581–1590, 2022. doi: 10.1016/j.procs.2023.01.136.
- [28] S. Panicacci, M. Donati, F. Profili, P. Francesconi, and L. Fanucci, "Trading-off machine learning algorithms towards data-driven administrative-socio-economic population health management", *Computers*, vol. 10, n° 1, pp. 1–21, 2021, doi: 10.3390/computers10010004.
- [29] D. R. S. Serrano, J. C. Rincón, J. Mejía-Resrepo, E. R. Núñez-Valdez, and V. García-Díaz, "Forecast of Medical Costs in Health Companies Using Models Based on Advanced Analytics", *Algorithms*, vol. 15, n° 4, 2022, doi: 10.3390/a15040106.
- [30] M. L. Hoffmann Souza, C. A. da Costa, and G. de Oliveira Ramos, "A machine-learning based data-oriented pipeline for Prognosis and Health Management Systems", *Computers in Industry*, vol. 148. Elsevier B.V., 2023. doi: 10.1016/j.compind.2023.103903.
- [31] N. Menteş, M. A. Çakmak, and M. E. Kurt, "Estimation of service length with the machine learning algorithms and neural networks for patients who receiving home health care," *Eval Program Plan*, vol. 100, 2023, doi: 10.1016/j.evalprogplan.2023.102324.
- [32] J. Dunstan et al., "Predicting no-show appointments in a pediatric hospital in Chile using machine learning", *Health Care Manag Sci*, vol. 26, n° 2, pp. 313–329, 2023, doi: 10.1007/s10729-022-09626-z.
- [33] G. Vos, K. Trinh, Z. Sarnyai, and M. Rahimi Azghadi, "Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review", *International Journal of Medical Informatics*, vol. 173, Elsevier Ireland Ltd, 2023. doi: 10.1016/j.ijmedinf.2023.105026.
- [34] V. C. Parro, M. L. M. Lafetá, F. Pait, F. B. Ipólito, and T. N. Toporcov, "Predicting COVID-19 in very large countries: The case of Brazil", *PLoS One*, vol. 16, n° 7, 2021, doi: 10.1371/journal.pone.0253146.