Cuaderno Activa
Revista científica de la Facultad de Ingeniería

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

Urrego, Johnny [1], Acosta, Juan [2]

Tipo de Artículo: revisión de literatura.

Recibido: 08/11/2023. Aprobado: 14/12/2024. Publicado: 00/00/0000

Resumen: Una revisión sistemática de literatura en la cual se busca comprender los términos, además de los estudios previos sobre Machine Learning (ML) y Bussines Intelligence (BI) para la predicción de una variable objetivo. La metodología utilizada incluyó una exhaustiva búsqueda de la literatura científica en bases de datos Scopus y ScienceDidirect y se seleccionaron estudios que cumplieran con criterios de inclusión predefinidos. Este artículo de revisión sistemática de literatura proporciona una visión general del ML y la IA aplicada y sus modelos. Los hallazgos destacan tanto los avances prometedores como los desafíos pendientes, lo que puede servir como 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 ScienceDidirect 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 future research and applications in the healthcare sector.

Keywords: Machine learning; predictive; health; growth; bussines intelligence.

PREPRINT

[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.urregopi@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

I. Introducción

Abordar la creciente necesidad de gestionar eficaz y eficientemente los recursos en el sector de la salud para aumentar la capacidad de prever con precisión cuántos pacientes se afiliarán a un plan de salud, puede ayudar a evitar la pérdida de recursos, optimizar la distribución de personal y proporcionar una mejor atención. Para esta RSL se hace indispensable conocer el comportamiento poblacional y predecir el aumento o disminución de una población, con ello pronosticar un posible gasto/inversión y por ende la rentabilidad. Estos modelos estadísticos estarían basados en modelos de *machine learning*. Para ello debe realizarse entonces la RSL clara, concisa y real sobre las metodologías y predicciones poblacionales, las posibles variables que puede requerir algún modelo a aplicar y la forma de implementarse.

Esta revisión busca conocer modelos predictivos que utilicen datos históricos y factores demográficos, entre otros, para ofrecer a las instituciones de salud una herramienta valiosa para la planificación y la toma de decisiones informadas, mejorando en última instancia la calidad de la atención médica, optimización de los recursos y la satisfacción de los pacientes; además nos explicaría con evidencia real y concreta las variables que pueden aplicar, las que pueden descartar, las formas y métodos estadísticos que deben implementarse y los datos necesarios para formular un modelo con un resultado claro, ajustado, relevante y creíble.

Se utilizó la búsqueda de publicaciones científicas en revistas indexadas en ScienceDirect y Scopus. La búsqueda se realizó en cualquier idioma y se centró en las motivaciones, las metodologías de trabajo, los desafíos y los beneficios que las organizaciones del sector salud han tenido al implementar la analítica de datos como una estrategia de gestión organizacional.

La inteligencia artificial (IA) y el aprendizaje automático (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 están empezando a influir en diversas áreas de la atención clínica, incluyendo la práctica médica.

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

II. Justificación de la revisión

Las técnicas de aprendizaje profundo y aprendizaje automático presentan oportunidades inigualables para mejorar la atención sanitaria, sin embargo, hay escasez de literatura sobre aplicaciones basadas en IA implementadas para mejorar la atención [2].

Los datos, la información y el conocimiento han experimentado un aumento exponencial en su impacto en la toma de decisiones, la implementación de nuevas estrategias y el crecimiento de las organizaciones. La ejecución de herramientas de aprendizaje automático (ML / IA) a nivel empresarial puede mejorar el registro, procesamiento, seguridad y el análisis de los datos para generar conocimiento. Esto puede permitir predecir y comprender los comportamientos de diversas variables y lograr una mejora continua en los procesos.

El uso de herramientas de ML puede incluir varios beneficios. Por ejemplo, puede ayudar a las organizaciones a identificar patrones y tendencias ocultas en grandes conjuntos de datos, lo que les contribuye a tomar decisiones más informadas y estratégicas. Además, el ML puede mejorar la eficiencia operativa al automatizar tareas repetitivas y optimizar procesos.

La implementación de herramientas de ML también puede contribuir a la seguridad y la protección de la información sensible. Al utilizar 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 interés gerencial y la capacitación del personal.

Existen proyectos en los cuales se programaban citas a través de un agente virtual *chatbot*, logrando ofrecer a los pacientes un nuevo canal de comunicación para la interacción del usuario. Es un claro ejemplo de la aplicabilidad de la inteligencia artificial y la *big data* en el área de la salud.

Es fundamental aclarar que se debe garantizar la privacidad y la ética en el manejo de los datos.

Con esta revisión sistemática de literatura, se quiere conocer cómo se han implementado herramientas de IA y de BI en la predicción poblacional de las EPS incluyendo diversas variables como condiciones socioeconómicas y si estas influyen en dichas predicciones.

PREPRINT

III. Formulación de las preguntas de investigación

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

1. ¿Qué variables se pueden utilizar 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 ofrecen resultados confiables, robustos y de alta calidad, para ayudar a validar y sostener las investigaciones realizadas, con mecanismos útiles y rápidos para la extracción, análisis y visualización de grandes volúmenes de datos, además de presentar la gran ventaja de ser aplicados en múltiples ramas del conocimiento desde economía, demografía, física, matemáticas y pesca, entre otras [3].

BI: Business Intelligence: es una de esas ideas y herramientas que combina datos operativos con herramientas analíticas para mostrar información compleja y competitiva a los planificadores y tomadores de decisiones. Las herramientas de inteligencia empresarial (BI) ayudan a las empresas a generar rápidamente conocimientos que guían a los gerentes hacia eficiencias operativas, los conducen a nuevas oportunidades y los diferencian de la competencia [4].

Big Data: es un término vago y genérico que puede abarcar varias propiedades distintas y no exclusivas. El origen del concepto de *Big Data* se atribuye a menudo a un breve informe técnico del grupo META, una consultora estadounidense y evolucionó hacia el concepto de 3V: Volumen – Variedad – Velocidad. Desde entonces se han propuesto otras V, como Veracidad. Todas las V se refieren a datos: los desafíos que enfrenta la sociedad de la información están ligados a grandes volúmenes de datos diversos y heterogéneos para procesar en tiempo real (Velocity) [5].

Crecimiento Poblacional: se refiere al cambio en el tamaño de una población durante un período de tiempo específico. Se puede medir como el cambio en el número de individuos en una población por unidad de tiempo [6].

EPS: según el artículo 177 de la Ley 100 de 1993 las define de la siguiente manera: las entidades promotoras de salud son las entidades responsables de la afiliación, y el registro de los afiliados y del recaudo de sus cotizaciones, por delegación del fondo de solidaridad y garantía [7].

IA: quizás la mayor revolución tecnológica del siglo XXI. Sus inicios se pueden situar a finales de los años 50, en donde la ciencia y la ingeniería se unen para fabricar máquinas inteligentes, especialmente programas informáticos inteligentes y 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 se definió ya en un artículo de Samuel de 1959 como un apartado de la IA que usa técnicas estadísticas y algoritmos computacionales para proporcionar a los ordenadores la capacidad de «aprender», es decir, mejorar sus resultados en una tarea específica tras procesar datos en cantidad y sin unas instrucciones explícitas externas (y por tanto potencialmente sesgadas) proporcionadas por el programador [9].

V. Proceso de búsqueda de documentos

En el proceso de la búsqueda de los documentos relacionados a la RSL, se realizó un análisis de artículos científicos en las bases de datos bibliográficas Scopus y ScienceDirect; en esta búsqueda no se restringió el idioma ni la ubicación y se seleccionaron los artículos de los últimos cuatro años que aplicaran con las palabras de búsqueda en inglés, por lo que se identificó qué idioma es el 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”, esto para ampliar el espectro de la búsqueda.

Posterior a la búsqueda inicial se procedieron a aplicar los filtros o “*limited to*” de año de publicación, tipo de documento, las temáticas, que el artículo fuera *open source* y un filtro muy importante que es el de “*keyword*” palabra clave en el cual el buscador brinda las opciones óptimas para afinar la búsqueda relacionada.

En la utilización de la “*keyword*” se encontraron términos adicionales para la búsqueda, que fueron, p*opulation statistics*, a*rtificial intelligence*, p*redictive value*, p*redictive analytics*, p*opulation research*, m*odels, statistical*, p*redictive models*, m*achine learning models*, p*rediction modelling* y p*rediction model*.

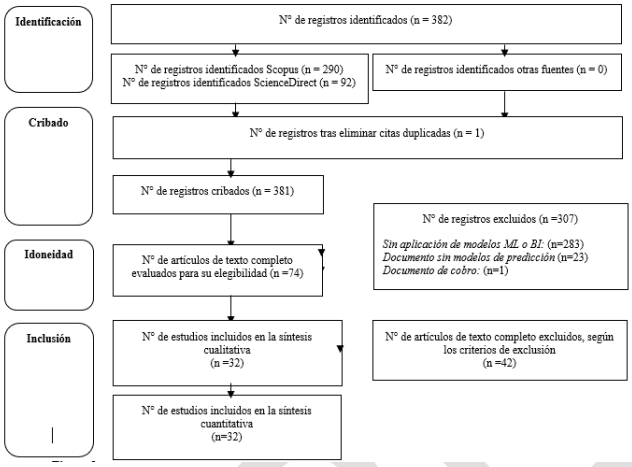
Arrojando 92 resultados en ScienceDirect y el *query* con los filtros de Scopus: 290.

PREPRINT

De los artículos encontrados con las palabras clave y los parámetros de búsqueda, se seleccionaron aquellos que estuvieran relacionados con el tema principal y cuya redacción fuera específica, clara y coherente. En este proceso, se realizó una lectura de los resúmenes para poder identificar los principales resultados y conclusiones que apuntan a la resolución de nuestro objetivo.

Se excluyeron documentos en los que no se hizo explícito los modelos de predicción, Además, de los artículos que eran netamente asistenciales, sin ninguna evidencia del uso de BI o de ML y finalmente aquellos en los que no se utilizó ninguna técnica de ML.

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], sobre un desarrollo de modelos de predicción de ausencias por enfermedad debida a trastornos mentales en la población activa general, investigó si las variables de las encuestas de salud ocupacional pueden usarse, y cómo, para identificar a los trabajadores en riesgo de sufrir una baja por enfermedad de larga duración debido a trastornos mentales. A partir de una consulta con un trabajador enfermo, se registra un código de diagnóstico derivado de la décima Clasificación Internacional de Enfermedades (CIE-10) y se realiza seguimiento por 1 año como variable de resultado.

Los autores Verma et al. [11] argumentan que no existe un marco claro e integral para la adopción de herramientas basadas en aprendizaje automático en salud, y proponen desarrollarlo en tres fases (exploración, diseño de soluciones e implementación y evaluación), sugiriendo utilizar un marco como CRISP DM7 para el desarrollo y prueba de algoritmos. Una gran ventaja de este enfoque es considerar la naturaleza iterativa de la 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 en Kampala, Uganda, Robsky et al. [12] evaluaron las asociaciones entre la movilidad, el acceso a la atención y la enfermedad de tuberculosis incluidos los viajes por motivos laborales, educativos o personales, teniendo en cuenta que estas variables pueden desempeñar un papel en la transmisión de esta enfermedad. Este estudio asoció la movilidad con un menor riesgo de enfermedad de tuberculosis, lo que puede deberse en parte al estatus socioeconómico más alto de las personas móviles en esta población.

Las variables que usaron para alimentar el modelo ML explotaron los cuatro parámetros climáticos: (temperatura, humedad, presión atmosférica y velocidad del viento) y dos parámetros de contaminación del aire (PM2,5 y PM10), los cuales se usaron como datos de entrada para predecir la infección y la recuperación. Estas pueden entenderse como las variables adicionales que pueden ayudar a enfocar y entrenar el modelo para una predicción más acertada [13].

Es importante conocer la estadística de los datos nacionales de indicadores demográficos y de salud, con el objetivo de crear modelos de predicción de la multimorbilidad, ejecutando, entre otros, un modelo de red neuronal de retro propagación. De esta manera, se pueden utilizar los predictores del modelo de regresión logística, tomando variables de género, hábito de fumar, edad y consumo de fruta que desempeñaron un papel importante en la predicción [14].

Los sistemas ciberfísicos (CPS) son sistemas que incorporan capacidad de computación y comunicación en su base para regular los procesos físicos y facilitar la interacción entre el mundo digital y el mundo físico en diversas tareas de control y monitoreo. En el ámbito de la salud, existe una variante de CPS conocida como Health CPS, la cual se utiliza como sistema de monitoreo de la salud. Este sistema tiene la capacidad de capturar, procesar y analizar de manera dinámica los datos provenientes de sensores de salud a través de procesos ciberfísicos integrados, los cuales están habilitados por el Internet de las cosas (IoT) [15].

PREPRINT

Los investigadores Mbunge y Batani [2], utilizaron múltiples sistemas o aplicaciones basadas en inteligencia artificial para mejorar la atención médica en el África subsahariana. Principalmente, utilizaron algoritmos de aprendizaje profundo y aprendizaje automático. Las técnicas de inteligencia artificial, particularmente el aprendizaje automático y los modelos de aprendizaje profundo, se han utilizado principalmente en diversas tareas, como predecir brotes de enfermedades, vigilar enfermedades, evaluar y diagnosticar riesgos, proporcionar atención personalizada en tiempo real, monitoreo remoto y atención preventiva.

En el artículo publicado por los autores 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 especifican variables a tener en cuenta para la aplicación de la inteligencia artificial en estudios de salud como atenciones, clinimetrías y fármacos. Además, hacen énfasis en diferentes repositorios de datos abiertos y privados para investigaciones en esta área; por ejemplo, la base de datos del laboratorio de Fisiología Computacional del Instituto Tecnológico de Massachusetts, El repositorio de códigos de MIMIC con más de 4.000 estudios que aprovechan los datos de registros médicos electrónicos de un hospital en Estados Unidos.

Hemos ingresado en una nueva era digital en el campo de la medicina, donde se vislumbra el potencial de aprovechar extensos volúmenes de información médica para agilizar descubrimientos científicos y elevar la calidad de la atención médica. Esto, a su vez, posibilita la atención médica personalizada y fundamenta la formulación de políticas basadas en evidencia. Este contexto, marcado por la convergencia de diversos factores, ha transformado los datos de atención médica, que incluyen información demográfica de pacientes, resultados de pruebas clínicas, hallazgos de laboratorio y datos genómicos, en un activo de gran valor tanto para entidades públicas como privadas. La creciente necesidad de vastas cantidades de datos de atención médica en ámbitos académicos y de la industria ha impulsado una lucrativa y enérgica tendencia hacia la comercialización, dando lugar a una industria multimillonaria que se enfoca en la recopilación, análisis y comercialización de dichos datos [17].

El empleo de análisis predictivo, apoyado en marcos basados en Internet de las cosas, ha generado un importante avance en los sistemas de salud inteligentes al permitir anticipar tendencias de mortalidad y, de esta manera, posibilitar intervenciones oportunas. Además, esta metodología facilita la identificación de relaciones ocultas entre variables relacionadas con la salud y factores críticos de mortalidad en el caso de niños. Para llevar a cabo este proceso, se emplearon conjuntos de datos provenientes de la Encuesta Demográfica de Salud de Etiopía y la Encuesta Demográfica de Salud de Pakistán. Estos conjuntos de datos del mundo real se sometieron a pruebas mediante el uso de clasificadores de aprendizaje automático, tales como Naïve Bayes, árbol de decisión, inducción de reglas, bosque aleatorio y perceptrón 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?

Algunos algoritmos que se han utilizado extensivamente para realizar modelos de aprendizaje automático han sido *Random Forest*, *Support Vector Machines* y Redes Neuronales. Adicionalmente, el *Deep Learning* o aprendizaje profundo se utiliza en aplicaciones más actuales. Este trabajo presentó una introducción al campo del aprendizaje automático, exhibiendo varias aplicaciones en salud con investigaciones que emplean el aprendizaje automático y logran resultados significativos en tareas que antes las realizaban exclusivamente humanos. Este hecho brinda nuevas oportunidades para mejorar la eficiencia de los servicios y prestaciones de salud, proporcionando un respaldo a la toma de decisiones clínicas [19].

El sistema de salud basado en IA mejora la predicción, el diagnóstico y las terapias de enfermedades, beneficiando tanto a los pacientes como a los proveedores de atención médica [8]. La aplicación de la IA en la atención sanitaria puede beneficiar a los médicos, pacientes y proveedores de atención sanitaria de cuatro maneras: I. estimando las posibilidades de éxito del tratamiento y analizando la aparición de la enfermedad antes del inicio del tratamiento; II. previniendo o manejando complicaciones; III. apoyando activamente la atención del paciente durante el diagnóstico y/o tratamiento y IV. identificando la patología de la enfermedad y también el mejor tratamiento [20].

Los servicios de salud móvil dotados de Inteligencia Artificial (IA) pueden conducir a mejoras importantes en términos de un seguimiento y una evaluación más automáticos, objetivos y continuos. Además, las últimas metodologías de IA explicable (XAI) pueden hacer que las decisiones de IA sean interpretables y confiables para los usuarios finales, esto nos lleva a que las decisiones tomadas basadas en modelos basados en IA y en este caso ML sean mucho más confiables y correctos [21].

En el estudio de Mirzaeian et al. [22] usaron en primer lugar la regresión logística binaria (BLR) para identificar los principales factores que influyen en el modelo. En segundo lugar, usaron ocho algoritmos de aprendizaje

PREPRINT

automático, que incluyen refuerzo adaptativo (*AdaBoost*), agregación de arranque (*Bagging*), refuerzo de gradiente extremo (*XG-Boost*), bosque aleatorio (RF), J-48, perceptrón multicapa (MLP), Naïve Bayes (NB), y la máquina de vectores de soporte (SVM) para predecir. Finalmente, su desempeño se evaluó utilizando métricas derivadas de la matriz de confusión para determinar el mejor modelo. Aunque son métodos y algoritmos que desconocemos, se utilizaron para la predicción de la esperanza de vida.

Los datos administrativos de salud longitudinales a nivel de la población sirven para facilitar el desarrollo de un sistema sanitario de aprendizaje. Desarrollando modelos de aprendizaje automático para predecir el riesgo de muerte entre los pacientes que llegan a un hospital. Es así como podemos inferir que la información es vital para la toma correcta de decisiones, enfoque de inversiones y por ende a la correcta gestión administrativa como asistencial tanto de las clínicas y hospitales como de aseguradoras de salud [23].

Aunque más del 80% de las veces, los modelos de ML sanitarios pueden predecir cómo se propagará una enfermedad, se buscará la forma de ajustar la predicción de comportamiento poblacional. En el experimento del documento mostraron que Random Forest predijo correctamente los casos positivos de malaria el 60 % de las veces y los casos positivos de tifoidea el 77 % de las veces, donde encontramos otro modelo para agregar a los posibles que podemos utilizar para predicción [24].

Con datos de características sociodemográficas, conductuales (inyección de drogas, comportamiento sexual) y biológicas para predecir los pacientes con alto riesgo de contraer VIH, el estudio de Nisa realiza análisis de características sociodemográficas, conductuales y biológicas para predecir los pacientes. Además, que informan el uso de minería de datos y aprendizaje automático, con lo cual se entiende qué tipo de variables podemos tener en cuenta para los ajustes de modelos [25].

En la metodología del aprendizaje automático en informática de la salud (SEMLHI) se incluyen cuatro módulos: (software, aprendizaje automático, algoritmos de aprendizaje automático y datos de informática sanitaria) que organizan las tareas en el marco utilizando una metodología SEMLHI, lo que permite a los investigadores y desarrolladores analizar software de informática sanitaria desde una perspectiva de ingeniería y proporciona a los desarrolladores una nueva hoja de ruta para el diseño de aplicaciones de salud con funciones de sistema e implementaciones de software [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 utilizando la inteligencia artificial, el *machine learning* y la inteligencia de negocios para mejorar el proceso de asignación de citas entre otras funcionalidades como cancelación de citas, revisión de historia clínica y confirmación al centro de salud correspondiente.

Un claro ejemplo de la aplicación de la *Big Data* y la inteligencia de negocios en el área de la salud se refleja en la investigación de Mohapatra et al. [27], donde usan teléfonos inteligentes y microsensores de signos vitales en dispositivos portátiles de monitoreo remoto. Allí sugirieron un marco para detectar cualquier situación de emergencia según la sensibilidad del modelo, junto con la transmisión oportuna de los datos del sensor y la recopilación y análisis de estos.

Una herramienta de gestión de la salud de la población basada en algoritmos inteligentes de última generación, a partir de datos administrativos y socioeconómicos, pronosticó que una cuarta parte de la población estadounidense tendrá más de 65 años en 2060. Esta situación no será diferente en los demás países. Además, se espera que la población de 85 años o más casi se duplique en los próximos 25 años. Los procesos automatizados de Big Data para definir cohortes de pacientes y para estratificar grupos por riesgo, buscan mejorar los resultados clínicos y reducir los costos, lo que apunta a una gestión óptima de las empresas de salud [28].

Un estudio realizado por el grupo Keralty Salud Colombia, sobre la previsión de costos médicos en empresas de salud utilizando modelos basados en analítica avanzada, aportó conocimiento desarrollando una estrategia para la previsión de costos y ayudó a optimizar los recursos y la toma de decisiones informadas mediante técnicas como un modelo estándar de memoria a largo plazo (LSTM) y agrupación de pacientes a través de *k-means clustering* [29].

La búsqueda de una utilización eficaz de los activos ha sido constante, especialmente en industrias con una mecanización en evolución. En este contexto, la gestión del mantenimiento gana visibilidad porque se encarga de garantizar la disponibilidad de los activos. El mantenimiento predictivo (PDM) es una de las principales estrategias de gestión del mantenimiento. Permite la detección temprana de fallas, evitando paradas no programadas y costos innecesarios. A medida que las tecnologías han avanzado, PDM ha evolucionado hasta convertirse en Gestión de pronóstico y salud (PHM), que proporciona los medios para reconocer patrones, comprender anomalías y estimar la vida útil restante (RUL) del equipo [30].

El autor Mentes [31], realizó un estudio sobre la estimación de la duración del servicio con algoritmos de aprendizaje automático y redes neuronales para

PREPRINT

pacientes que reciben atención médica domiciliaria, realizado en 14 hospitales que brindan servicios de atención médica domiciliaria en la ciudad de Diyarbakır, Turquía. Los datos obtenidos fueron sometidos a un pre-procesamiento necesario y se aplicó estadística descriptiva. Se aplicaron varios modelos para encontrar el más ajustado: árbol de decisión, bosque aleatorio y red neuronal de perceptrón multicapa con lo que se encontró que el número de días de servicio de atención médica domiciliaria que recibieron los pacientes varió dependiendo de su edad y sexo. Se observó que los pacientes generalmente se encontraban en los grupos de enfermedades que requerían tratamientos de Fisioterapia y Rehabilitación. Se determinó que la duración del servicio de los pacientes se puede predecir con una alta tasa de confiabilidad (Modelo multicapa: 90,4%, Modelo de árbol de decisión: 86,4% y Modelo de bosque aleatorio: 88,5%) utilizando algoritmos de aprendizaje automático [31].

Con el ánimo de predecir citas de ausencia en un hospital pediátrico de Chile mediante aprendizaje automático, un grupo de investigadores diseñaron una plataforma computacional para implementar los modelos predictivos como una aplicación web. El *front-end* y el *back-end* fueron diseñados en Python usando el *framework web* Django. La entrada es una hoja de cálculo que contiene las características de la cita, como la identificación del paciente y otra información personal, especialidad médica, fecha y hora. Estos datos se procesan para generar las características descritas para cada especialidad, las etiquetas de todas las citas se predicen utilizando el mejor modelo predictivo. Las citas se clasifican en orden descendente según la probabilidad prevista de no presentarse, junto con la información de contacto del paciente. Luego, el hospital puede comunicarse con los pacientes con mayor probabilidad de no presentarse para confirmar la cita [32].

Los sensores portátiles se han mostrado prometedores como método no intrusivo para recopilar biomarcadores que pueden correlacionarse con niveles elevados de estrés. Los factores estresantes causan una variedad de respuestas biológicas, y estas reacciones fisiológicas se pueden medir utilizando biomarcadores, cambios en las condiciones de salud que pueden correlacionarse con niveles elevados de algún factor, y con ello ayudar a mejorar la prevención de la enfermedad, que en últimas puede traducirse en reducción de costos para las instituciones. El seguimiento y monitoreo de la salud mediante dispositivos portátiles está ganando popularidad, mientras que la generalización de los modelos de aprendizaje automático existentes aún requiere más estudios, y la investigación en esta área continuará brindando mejoras a medida que se disponga de conjuntos de datos más nuevos y sustanciales [33].

El trabajo de Parro [34] y otros, presenta una propuesta práctica para estimar la utilización del sistema de salud para casos de COVID-19. La novedosa metodología desarrollada se basa en el modelo dinámico conocido como Susceptibles, Infectados, Eliminados y Muertos (SIRD). El modelo se modificó para centrarse en la dinámica del sistema sanitario, en lugar de modelar todos los casos de la enfermedad. Se ajustó utilizando datos disponibles para cada estado brasileño y se actualizó con cifras diarias. Se definió y utilizó una figura de mérito que evalúa la calidad del ajuste del modelo a los datos para optimizar los parámetros libres. Se estimaron los parámetros de un modelo epidemiológico para todo Brasil, que comprende una combinación lineal de los modelos para cada estado, considerando los datos disponibles para los 26 estados brasileños. [34] el modelo se validó y se demostró una fuerte adherencia en la mayoría de los casos.

VII. Conclusiones

En el ámbito de salud, los beneficios de las soluciones basadas en inteligencia artificial son tan grandes que cada vez más se suman empresas aseguradoras y prestadores que están interesados en hacer inversión en IA, lo que demuestra que también existe una dimensión comercial, además de que la demanda de estas soluciones aumenta su importancia y valor en el sector sanitario.

Dado que existen diversos proyectos de desarrollo en el campo de la IA que pueden desarrollar proyectos de manera significativa y agregar valor a las EPS, el ML está aportando un gran potencial al sector de la salud. Estos proyectos ayudan a mejorar la eficiencia y precisión de los diagnósticos médicos, optimizar los procesos de atención al paciente y facilitan la toma de decisiones administrativas como clínicas. La implementación de soluciones basadas en ML puede proporcionar resultados más rápidos y precisos, lo que potencialmente conducirá a una mejor calidad de la atención y menores costos. Además, estos proyectos pueden ayudar a identificar patrones y tendencias en grandes conjuntos de datos asistenciales, contribuyendo a la investigación y el desarrollo de nuevos programas y/o tratamientos, a mejorar la calidad de la atención y permitir una toma de decisiones más informada y precisa.

Hoy en día, la combinación de IA (Inteligencia Artificial), *Big Data*, ML (aprendizaje automático) y BI (*Bussines Intelligence*) es esencial para optimizar los recursos y maximizar los resultados. Estas tecnologías trabajan juntas para recopilar, procesar y analizar grandes cantidades de datos, lo que permite obtener información valiosa y tomar decisiones más informadas. La IA y el

PREPRINT

ML pueden ayudar a identificar patrones y tendencias ocultos en los datos y mejorar la eficiencia y eficacia de los procesos comerciales. *Big data* y BI, por otro lado, proporcionan las herramientas necesarias para almacenar, gestionar y visualizar datos de forma eficiente y eficaz. Al utilizar estas tecnologías las empresas pueden obtener una ventaja competitiva al optimizar los recursos y lograr mejores resultados en los procesos de negocio.

La continuidad del desarrollo es un tema muy versátil, diverso, y sus beneficios apenas comienzan a explorarse en profundidad. A medida que avanza la tecnología, se abren nuevas oportunidades en el ámbito del desarrollo continuo, permitiendo la mejora continua de productos y servicios. Al adoptar metodologías ágiles y enfoques de desarrollo iterativos, las empresas pueden adaptarse rápidamente al cambio y ofrecer soluciones más eficientes. Aunque todavía queda mucho por descubrir y desarrollar en este ámbito, está claro que futuros avances pueden tener un impacto significativo en diferentes sectores y áreas de negocio. Profundizar en los beneficios puede revelar nuevas oportunidades para optimizar los procesos, mejorar la experiencia del usuario y aumentar la eficiencia, eficacia y efectividad.

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, "Ley100 1993".

[8] R. B. Ruiz y J. D. Velásquez, "Artificial intelligence at the service of the health of the future", Revista Medica Clinica Las Condes, vol. 34, n° 1, pp. 84–91, 2023, doi: 10.1016/j.rmclc.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–E1357, 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.

PREPRINT

[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 Medica Clinica Las Condes, vol. 33, n° 6, pp. 568–575, 2022, doi: 10.1016/j.rmclc.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] 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-demographical, 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 Appointmentsin 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-Restrepo, 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 Plann, 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.

PREPRINT