

SISTEMAS

$$n(B) = 68$$

$$n(C) = 84$$

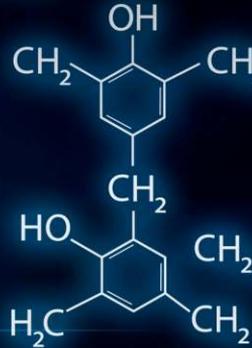
$$n(B \cup C) = n(B) + n(C) - n(B \cap C)$$

$$\text{He} = 4.002602$$

$$\text{Na} = 22.989769$$

$$\text{Ar} = 39.948$$

$$M = \frac{0.046765 \text{ mol}}{3 \text{ OL}} = 0$$

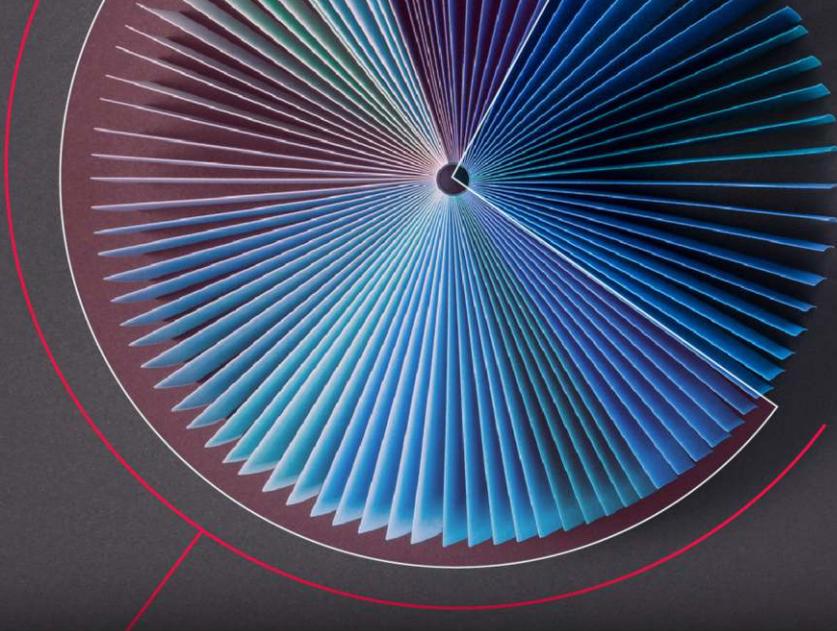


MLOps (Machine Learning Dev Ops)



ASOCIACIÓN COLOMBIANA DE INGENIEROS DE SISTEMAS

Calle 93 No. 13 - 32 of. 102
Bogotá, D.C.
www.acis.org.co



Sabías que **los datos más estratégicos de tu negocio son ignorados** a pesar de que tienes acceso a ellos



Somos **especialistas con más de 20 años de experiencia** en manejo técnico y estratégico de datos. Hablemos y **transformemos el futuro de sus resultados.**

www.gpstrategy.com.co
hablemos@gpstrategy.com.co

GPStrategy
by  **MERCANZA**

En esta edición

Editorial

4

Cadena de suministro digital

DOI: 10.29236/sistemas.n165a1

La nueva frontera de la resiliencia de las empresas

Columnista Invitado

8

Machine Learning y DataOps

DOI: 10.29236/sistemas.n165a2

La automatización de los procesos de Machine Learning facilitan el desarrollo continuo de Analítica

Entrevista

12

Desafíos de la humanidad

DOI: 10.29236/sistemas.n165a3

Luis Murillo Fernández se pronuncia sobre “Machine Learning” con sus 15 años de experiencia en inteligencia artificial.

Investigación

16

Machine Learning Operations y sus desafíos de implementación en Colombia

DOI: 10.29236/sistemas.n165a4

El estudio sobre tecnologías inteligentes, realizado por ACIS a través de Internet, recibió respuesta de ochenta encuestados a nivel nacional.

Cara y Sello

23

Machine Learning Operations

DOI: 10.29236/sistemas.n165a5

Los desafíos de su implementación en Colombia.

Uno

30

Machine Learning

DOI: 10.29236/sistemas.n165a6

Como apoyo a la medicina predictiva.

Publicación de la Asociación Colombiana de
Ingenieros de Sistemas (ACIS)
Resolución No. 003983 del
Ministerio de Gobierno
Tarifa Postal Reducida Servicios Postales
Nacional S.A. No. 2015-186 4-72
ISSN 0120-5919
Apartado Aéreo No. 94334
Bogotá D.C., Colombia

Dirección General

Jeimy J. Cano M.

Consejo de Redacción

Francisco Rueda F.
Gabriela Sánchez A.
Manuel Dávila S.
Andrés Ricardo Almanza J.
Emir Hernando Pernet C.
Fabio Augusto González O.
Jorge Eliécer Camargo M.
María Mercedes Corral S.

Editor Técnico

Joseph Alejandro Gallego M.

Coeditor Técnico

Fabio Augusto González O.

Editora

Sara Gallardo M.

Junta Directiva ACIS

2022-2024

Presidente

Luis Javier Parra B.

Vicepresidente

Jorge Fernando Bejarano L.

Secretario

Rodrigo Rebolledo M.

Tesorero

Jaime García C.

Vocales

Hilda Cristina Chaparro L.
Soledad Mercedes Gutiérrez R.

Directora Ejecutiva

Beatriz E. Caicedo R.

Diseño y diagramación

Bruce Garavito

Los artículos que aparecen en esta edición no
reflejan necesariamente el pensamiento de la
Asociación. Se publican bajo la responsabilidad
de los autores.

Octubre - Diciembre 2022

Calle 93 No.13 - 32 Of. 102
Teléfonos 616 1407 - 616 1409
A.A. 94334
Bogotá D.C.
www.acis.org.co

NASCO

NACIONAL DE COMPUTADORES S.A.

APOYA ESTA PUBLICACIÓN

TEL: 6 06 06 06- CR 15 No 72-73



Confía en 4-72,
el servicio de envíos
de Colombia

Línea de atención al cliente:
(57 - 1) 472 2000 en Bogotá
01 8000 111 210 a nivel Nacional

.....
www.4-72.com.co

AFFILIATE

BENEFICIOS

- Inclusión en el gremio de Ingenieros de Sistemas más importante del país.
- Actualización y formación profesional y académica de manera constante.
- Descuentos especiales en cursos y eventos exclusivos en el área de las TIC.
- Recibirán la Revista Sistemas (Publicación virtual Trimestral).
- Pertener a los grupos de interés especializados en el sector.
- Participación activa como conferencista en ACIS.
- Oportunidad laboral en nuestros portales Oferta de Empleo y Perfil de Ingenieros
- Los Asociados pueden asistir todos los JUEVES a las conferencias de forma gratuita. Consulte la Programación de Conferencias .
- Asista a todas las funciones programadas por el Teatro Nacional con un 20% de descuento. Consulte toda la Programación del Teatro Nacional y solicite el descuento a cursos@acis.org.co.
- Adquiera un 30% de descuento en los libros de la Casa Editorial ALFAOMEGA, consulte el Catálogo

MAS INFORMACION EN :
WWW.ACIS.ORG.CO
3015530540 - 3043463413

Machine Learning Operations

DOI: 10.29236/sistemas.n165a1



Desafíos de su implementación en Colombia.

Joseph Alejandro Gallego Mejía

El análisis de datos y el aprendizaje de máquina están modificando cada aspecto actual de nuestras vidas. Actividades que antes eran realizadas enteramente por seres humanos como la detección de enfermedades, la detección de objetos, el reconocimiento y sínte-

sis de voz, entre otros, hoy en día están siendo automatizadas usando algoritmos de aprendizaje de máquina.

El aprendizaje profundo (en inglés *deep learning*) lleva más de una década mejorando los resultados de

métodos clásicos de aprendizaje de máquina, superando a los humanos en varios campos. Como ejemplo tenemos a AlphaZero creado en DeepMind de Google que nos ha aventajado en varios juegos, como el ajedrez y go.

De la misma forma como en los noventa había una necesidad de mejorar el ciclo de vida del desarrollo de software, hoy en día necesitamos mejorar el de los modelos de aprendizaje de máquina. Estos problemas están relacionados con la creación, el empaquetamiento, el despliegue, las pruebas, y la posterior monitorización del modelo.

Por lo tanto, es necesario crear buenas prácticas en el ámbito del aprendizaje de máquina para evitar la deuda técnica generada al usar soluciones simples en problemas complejos.

Enmarcados en este contexto, el comité editorial de la Revista Sistemas decidió dedicar este número a discutir diversos aspectos del MLOps. Se abordan las definiciones de aprendizaje de máquina, sus distintas ramas, su uso en aplicaciones médicas y una discusión acerca de su utilización dentro de las empresas colombianas

En el artículo uno, el ingeniero Holman D. Bolívar B. hace un amplio recorrido desde la construcción de herramientas hasta algoritmos de inteligencia artificial realizados por el ser humano. Finaliza su análisis

con la definición y explicación de metodologías ágiles, integración y despliegue continuo, además de Machine Learning Operations.

En el artículo dos, los ingenieros Luis A. Blanquicett B. y Luis F. Murillo F. abordan el uso del aprendizaje de máquina en la medicina.

Los autores abordan la definición de los tipos de aprendizaje de máquina y luego explican en detalle los usos del aprendizaje de máquina en medicina.

Para la entrevista tuvimos como invitado al ingeniero Luis Murillo Fernández quien se pronunció sobre “Machine Learning” desde sus quince años de experiencia en inteligencia artificial. Inicia su intervención diciendo que hoy en día tenemos grandes avances en la inteligencia artificial en general, con grandes retos éticos y un faltante en la discusión sobre la conciencia biológica que hoy conocemos.

La sección de investigación se basó en una encuesta nacional realizada por la Asociación Colombiana de Ingenieros de Sistemas (Acis), en la que se indaga por el conocimiento de MLOps, así como por los distintos roles de análisis de datos y el uso de estas herramientas.

En la sección de Cara y Sello participaron varios expertos nacionales de la academia y la industria, quienes compartieron sus opiniones alrededor de diversos aspectos de

MLOps y el aprendizaje de máquina. Se discutió sobre los principales riesgos en los proyectos de Machine Learning que desencadenan en deuda técnica, el desafío de llevar modelos creados en cuadernos interactivos a producción, los desafíos técnicos del ciclo de vida de modelos de aprendizaje de máquina, además de los retos y posibles soluciones que tiene Colombia concerniente a MLOps.

Como se puede apreciar, este número recoge varias perspectivas y opiniones alrededor de un tema que está en pleno desarrollo. De ahí que contemple una mirada incompleta, por tratarse de un tema vasto y en crecimiento.

No dudamos que nuestros lectores encontrarán un factor motivador para profundizar más en esta apasionante área. 🌐

Joseph Alejandro Gallego Mejía. Candidato a doctor en Ingeniería de Sistemas y Computación. Magíster en Ingeniería de Sistemas e Ingeniero de Sistemas y Computación de la Universidad Nacional de Colombia. Ingeniero Industrial de la Universidad Nacional de Colombia. Investigador en el área de las ciencias de la computación. Gerente de tecnología y fundador de Sammu, empresa internacional con presencia en Colombia, Chile, y Perú. Desarrollador web, móvil y back-end. Docente de la Universidad Nacional de Colombia, del programa Misión TIC 2022. Investigador asociado al grupo MindLab de la Universidad Nacional de Colombia, sede Bogotá; realiza investigación en aprendizaje de máquina, aprendizaje estadístico, minería de datos y análisis de datos.



GEO DATOS **2023**

**SERVICIOS PÚBLICOS Y
TRANSFORMACIÓN ENERGÉTICA
ALINEADOS CON LAS TIC Y LOS SIG**

**MAS INFORMACION EN :
WWW.ACIS.ORG.CO
3015530540 - 3043463413**

Machine Learning y DataOps

DOI: 10.29236/sistemas.n165a2



La automatización de los procesos de Machine Learning facilitan el desarrollo continuo de Analítica.

Holman Diego Bolívar Barón

Desde tiempos inmemoriales, los seres humanos han construido máquinas para simplificar su trabajo y reducir el esfuerzo en completar diferentes tareas. Sin conocer ninguna de las leyes de la mecánica Newtoniana o los teoremas mecá-

nicos de Arquímedes, inventaron palancas, instrumentos y máquinas complejas para realizar procedimientos sofisticados. Mover grandes y pesadas rocas o troncos de madera se hizo más sencillo utilizando un carro. Los pueblos primi-

tivos simplemente observaron cómo la rueda podía mejorar sus vidas (Ethem, 2016).

Una máquina se considera útil si sus usuarios pueden entender fácilmente qué tareas se pueden completar con menos esfuerzo o automáticamente, máquinas automáticas, construidas y programadas para lograr objetivos específicos al transformar la energía en trabajo, los molinos de viento son un ejemplo de herramientas elementales que pueden llevar a cabo tareas completas con un control humano mínimo.

Los seres humanos han tratado de transferir algo de inteligencia a sus herramientas desde los albores de la tecnología. Una rueda puede girar alrededor de un eje fijo millones de veces y el viento debe encontrar una superficie adecuada para empujarla, el camino comenzó con las herramientas, prosiguió con las máquinas simples y luego a las inteligentes (Rahman, 2020).

Las computadoras programables son instrumentos muy extendidos, flexibles y cada vez más poderosos; además, la difusión de Internet permitió compartir aplicaciones de software e información relacionada con un mínimo de esfuerzo. El software de procesamiento de texto, un cliente de correo electrónico, un navegador web y muchas otras herramientas comunes que se ejecutan en la misma máquina, son ejemplos de esa flexibilidad. Es innega-

ble que la revolución de TI mejoró nuestros trabajos diarios, pero sin el aprendizaje automático ML (por Machine Learning), todavía habría muchas tareas que serían ajenas a la informática. El filtrado de spam, el procesamiento de lenguaje natural NLP (por Natural Language Processing), el seguimiento visual con una cámara web o un teléfono inteligente y el análisis predictivo son solo algunas de las aplicaciones que revolucionaron la interacción hombre-máquina, transformando herramientas electrónicas en extensiones cognitivas reales que están cambiando y reduciendo la brecha entre la percepción humana, el lenguaje, el razonamiento y entidades artificiales.

ML no se basa en estructuras estáticas o permanentes, sino en una capacidad continua para adaptar su comportamiento a señales externas correspondientes a conjuntos de datos o entradas en tiempo real y, como un ser humano, estimar un escenario o situación futura con información incierta y fragmentaria (Bonaccorso, 2018).

Por otra parte, la cantidad de datos que se genera es inconmensurable, con base en la publicación de Lori Lewis (2021) en el año 2021 durante 60 segundos se realizaban 1.4 millones de scrolling en Facebook, 9.132 conexiones a LinkedIn, 28.000 suscriptores de Netflix estaban viendo alguna serie o película, 1.6 millones de personas estaban haciendo compras online, 69

millones de personas enviaban mensajes por WhatsApp, se generaron 200.000 twits y 500 horas de video eran subidas y reproducidas en Youtube, entre muchas otras.

Actualmente, la cantidad de datos que se producen, se transportan y se transforman es superior a la cantidad de datos generados por la humanidad a lo largo de la historia, por tal razón, no es posible leer, comparar y analizar todos los datos, incluso para supercomputadoras esto puede llegar a tomar mucho tiempo, por tal razón es necesario recurrir a técnicas estadísticas para el análisis de la gran explosión de datos que se genera cada minuto.

Todo modelo de ML requiere mínimo un dataset de entrada, sin embargo, para implementar el modelo, este requiere cierta homogeneidad entre los datos para compararlos e identificar patrones y estructuras que permitan establecer conjuntos o grupos de datos. Un grupo homogéneo de datos es lo que se conoce comúnmente como clúster.

El agrupamiento es el proceso de análisis para clasificar datos en una serie de grupos, en cada grupo los datos son similares entre sí bajo ciertos parámetros o criterios, entre más criterios cumple un grupo de datos se establece que la calidad del grupo es mayor. El agrupamiento es comúnmente denominado clustering (por su denominación en inglés). En clustering tam-

bién es posible segmentar clusters con propiedades o comportamiento similares y asignar los datos a alguno de ellos. Es posible que no se logre identificar un parámetro o criterio para realizar clustering con los datos, pero es necesario una segmentación, por lo tanto, los datos se pueden agrupar aleatoriamente, esta técnica es conocida como CLARANS (por Clustering Large Applications based upon Randomized Search).

El proceso de calidad de los datos es fundamental dentro de un ejercicio de ML, para ello se suelen desarrollar ejercicios de ETL (por Extract, Transform and Load) que homogenizan los datos para que sean comparables y evaluables entre sí. Una sola proyección puede llegar a requerir la integración de decenas de sistemas de información para estructurar ese Dataset que permite realizar el ejercicio de ML.

Ese proceso de homogenización se puede desarrollar realizando ejercicios de ingeniería de datos apoyándose en nuestro bien apreciado y ponderado SQL (por Structured Query Language) también se puede desarrollar con funciones en lenguaje R o Python. Así mismo, existen herramientas para automatizar este proceso como Databricks, Stratio o Alteryx.

En proyectos robustos de ML se debe asegurar el ciclo de vida del dato y el cumplimiento de los lineamientos generales para la imple-

mentación de la infraestructura de datos con base en la resolución 460 del 15 de febrero de 2022 del Ministerio de Tecnologías de la Información y las Comunicaciones de Colombia.

Para poder abordar la complejidad, incertidumbre y ambigüedad en el desarrollo de los productos de ML y teniendo en cuenta el artículo 2.2. 9.1.1.3 del Decreto 767 del 16 de mayo de 2022 y el lineamiento del documento Maestro del Modelo de Gestión de Proyectos TI, planteado por el Ministerio de Tecnologías de la Información y Comunicaciones, el cual establece en el numeral 7.2. 10. MGPTI.LI.PLA.10 que: “Se debe considerar el uso de metodologías ágiles y la aplicación del manifiesto Ágil”.

Una metodología ágil podría ser Scrum, la cual permite el trabajo colaborativo mientras se aborda cada uno de los problemas generados durante el desarrollo e implementación, “Con scrum, un producto se basa en una serie de iteraciones llamadas sprints que dividen proyectos grandes y complejos en porciones minúsculas”, “Gracias a los sprints, los proyectos son más fáciles de gestionar, permiten a los equipos enviar trabajo de gran calidad más rápido y con más frecuencia, y les ofrecen más flexibilidad para adaptarse al cambio” (Drummond, 2022).

Scrum como metodología de desarrollo de productos de software

se enfoca en un desarrollo continuo lo que permite generar innovación, pasando por diferentes MVP (por Minimum Viable Product); sin embargo, la complejidad del despliegue dependerá de los recursos para la correcta funcionalidad de la aplicación, así como la interacción con el o los sistemas operativos sobre los cuales se ejecutan los servidores que soportan la funcionalidad de esta. Cuando la aplicación se mantiene en evolución y desarrollo constante, el despliegue debe automatizarse bajo el paradigma de integración y despliegue continuo CI/CD (por Continuous Integration and Continuous Delivery), lo que actualmente se denomina DevOps (por Development and Operations).

En el marco de los ejercicios de analítica de datos se sigue el esquema propuesto por IBM a través del CRISP-DM (Por Cross Industry Standard Process for Data Mining) que contempla una fase de entendimiento del negocio y de los datos, seguido de fases de preparación, modelamiento, evaluación e implementación.

A la hora de implementar una metodología que permita la automatización del despliegue de ejercicios de ML aparece MLOps (por Machine Learning Operations), ya que proporciona un proceso de desarrollo de ML de extremo a extremo para diseñar, construir y administrar software reproducible, comprobable y evolutivo impulsado por ML.

MLOps tiene como objetivo unificar el ciclo de lanzamiento para el aprendizaje automático y el lanzamiento de aplicaciones de software, además, permite hacer pruebas automatizadas de artefactos de aprendizaje automático (por ejemplo, validación de datos, pruebas de modelos de ML y pruebas de integración de modelos de ML). MLOps facilita la aplicación de principios ágiles a proyectos de aprendizaje automático y reduce la deuda técnica en todos los modelos de aprendizaje automático. MLOps debe ser una práctica independiente del lenguaje, el marco, la plataforma y la infraestructura y permite admitir modelos y conjuntos de datos de aprendizaje automático para construir estos modelos como ciudadanos de primera clase dentro de los sistemas de CI/CD.

Referencias

Bonaccorso. Giuseppe., (2018). Machine Learning Algorithms: Popular Algorithms

for Data Science and Machine Learning, 2nd Edition: Vol. 2nd ed. Packt Publishing. Drumond. Claire, (2022), What is scrum?, Atlassian, disponible en: <https://www.atlassian.com/agile/scrum>

Ethem Alpaydin. (2016). Machine Learning : The New AI. The MIT Press.

IBM, (2021), CRISP-DM Help Overview, disponible en: <https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=dm-crisp-help-overview>

Lewis. Lori, (2021), 2021 This is What Happens in An Internet Minute. <https://www.allaccess.com/merge/archive/32972/infographic-what-happens-in-an-internet-minute>.

MinTIC, Artículo 2.2.9.1.2.1. “Componentes” del Capítulo 1 “ESTRATEGIA DE GOBIERNO EN LÍNEA” del Título 9 del Decreto 1078 de 2015

Rahman. Was., (2020). AI and Machine Learning. Sage Publications Pvt. Ltd. 

Holman Diego Bolívar Barón. Doctor en Informática por la Universidad Pontificia de Salamanca, miembro profesional de la Association for Computing Machinery (ACM), líder del grupo de investigación en Software Inteligente y Convergencia Tecnológica –GISIC-. DEA en Ingeniería del Software, profesor de la Universidad Católica de Colombia e investigador asociado de Minciencias. Su área de investigación se centra en el desarrollo de software para entrenamiento cognitivo y modelos de desarrollo para la evaluación del aprendizaje en entornos interactivos a través de la lógica difusa. Actualmente trabaja en la estructuración de una plataforma de ciencia de datos y aprendizaje de máquina automatizado de código abierto para la investigación de acceso público al servicio de la sociedad.

Calendario de Eventos

2023

MAR
06 - 10

GEODATOS

MARATÓN REGIONAL LATINOAMERICANA
DE PROGRAMACIÓN ACIS/REDIS

MAR

ABR

JORNADA DE GERENCIA DE PROYECTOS TI

JORNADA DE CIUDADES + TECNOLOGÍA

MAY

JUN

JORNADA INTERNACIONAL DE SEGURIDAD
INFORMÁTICA

ACISTIC

AGO

SEP

MARATÓN NACIONAL DE PROGRAMA-
CIÓN ACIS/REDIS

REDUC@TE

OCT

NOV

ENCUENTRO DE REDIS



MAS INFORMACION EN :
WWW.ACIS.ORG.CO
3015530540 - 3043463413

Desafíos de la humanidad

DOI: 10.29236/sistemas.n165a3

Luis Murillo Fernández se pronuncia sobre “Machine Learning” con sus 15 años de experiencia en inteligencia artificial.

Sara Gallardo M.

Luis Murillo Fernández es ingeniero electricista con especialización en Informática Industrial y maestría en Gestión de la innovación. Su experiencia en Inteligencia Artificial es de 15 años, “inicialmente con sistemas de control neuro difuso y hoy en día en Machine Learning y ciencia de datos”, aclara.

Se define como un apasionado por su trabajo y señala que el tiempo libre lo dedica a su familia y a sus hobbies: la electrónica y las manualidades con madera. Le encantan las películas de ciencia ficción.

“Estamos en un hito importante; por un lado, los avances de la inteligencia artificial en general, sus retos éticos y las discusiones sobre la conciencia biológica que hoy cono-

ceamos. El mundo se enfrenta a desafíos importantes desde la igualdad, la guerra y la carrera armamentista; además, a los nacionalismos que influyen en las decisiones tomadas hoy y que tendrán un impacto significativo en el futuro de la humanidad”, no duda en afirmar, antes de entrar de lleno en la entrevista.

RS: Desde su experiencia, ¿cómo define el muro de la confusión en proyectos de Machine Learning y cómo MLOps ayuda a tumbar ese muro?

Luis Murillo Fernández: El llamado muro de la confusión es un concepto que va más allá de la organización en la forma como se estructura y también en las personas que conforman los correspondientes



equipos de trabajo. Nos encontramos con este marco de objetivos diferentes entre los distintos departamentos. Debido a lo novedoso de los desarrollos en Machine Learning, la diferencia de objetivos entre los equipos de desarrollo y de operación es mucho más amplia; los primeros intentando crear nuevas y mejoradas funcionalidades y los segundos intentando tener sistemas más estables. MLOps como proceso de múltiples funciones permite implementaciones rápidas, colaborativas e iterativas que ponen en marcha las capacidades de

Machine Learning y la Ciencia de Datos, reduciendo el riesgo de realizar proyectos en desarrollo que no llegan a producción y operación.

RS: *¿Cuáles son los principales riesgos en los proyectos de Machine Learning que desencadenan en deuda técnica?*

LMF: Esta deuda técnica en Machine Learning puede aparecer por diversos factores, entre los más comunes está la baja calidad de los datos utilizados para el entrenamiento de los modelos (overfitting y underfitting), lo cual puede llevar a

no elegir el mejor por el afán de desplegar la solución.

RS: *Muchos científicos de datos desarrollan buena parte de su trabajo usando herramientas interactivas como los cuadernos de Jupyter. ¿Cómo pueden estos profesionales evolucionar su trabajo hacia el uso de MLOps?, ¿en qué aspectos se deberían enfocar primero?*

LMF: El uso de MLOps por parte de los científicos de datos es un asunto pendiente que aún tenemos muchos al desarrollar soluciones de Machine Learning y ciencia de datos. Es importante que durante el entrenamiento y capacitación de estas personas se incluya el uso de herramientas de MLOps, que en muchas de las formaciones no se encuentran o se describen muy superficialmente.

RS: *¿Cuáles son los desafíos técnicos a los cuales se enfrenta un equipo de desarrollo al momento de iniciar con MLOps?*

LMF: Los desafíos técnicos al iniciar con MLOps están relacionados más con los conocimientos y la experiencia del equipo, que con las herramientas y plataformas adecuadas. Pero indudablemente existen otros factores que pueden presentar desafíos tales como no disponer por parte de la organización de los recursos necesarios, no dar tiempo para que los equipos de trabajo puedan experimentar con MLOps antes de abordar proyectos formales; no planificar adecuadamente

los proyectos y no definir claramente los resultados esperados, los inconvenientes y contingencias que pudieran encontrarse en su desarrollo. Y, no menos importante, es que las personas y organizaciones muchas veces tienen problemas con los cambios y MLOps implica necesariamente cambiar la forma como se hacen los proyectos de Machine Learning en la organización.

RS: *¿Son los datos tan importantes como lo plantea Andrew NG en un proyecto de machine learning?*

LMF: La calidad de los datos utilizados en Machine Learning y ciencia de datos es muy importante, toda vez que estos son la materia prima con la que entrenaremos los modelos y por tanto tienen un efecto en la precisión y escalabilidad futura de nuestros proyectos, a diferencia del desarrollo de software tradicional, en el que las mejoras van de la mano para optimizar la arquitectura o el código.

RS: *¿Cómo se articulan los datos dentro de MLOps?*

LMF: En MLOps los datos siguen siendo importantes como en cualquiera de nuestros sistemas basados en Machine Learning, pero la automatización de procesos analíticos (APA), permite estandarizar y acelerar las primeras etapas de acceso a los datos, su preparación, el modelado y ajuste, que en proyectos tradicionales de Machine Learning y ciencia de datos podrían requerir más tiempo.

RS: *¿Está Colombia preparada para asumir los retos de MLOps?*

LMF: En gran medida sí, pero sigue existiendo un déficit de personas capacitadas en Machine Learning y ciencia de datos, y más aún con conocimientos y experiencia en MLOps.

RS: *¿Cómo podríamos mejorar las habilidades técnicas de las personas involucradas en MLOps en Colombia?*

LMF: Siempre es importante la capacitación y el entrenamiento, participar de proyectos reales que involucren MLOps para lograr que las personas logren mejorar ampliamente sus habilidades técnicas.

RS: *¿Debería ser una prioridad del gobierno no sólo preparar nuevos científicos de datos sino personas expertas en MLOps? ¿Por qué? ¿Con qué objetivo?*

LMF: Si, desde hace algunos años

se han estado desarrollando convocatorias para capacitación que han dado buenos resultados, pero en muchas de ellas, MLOps no ha estado presente para mejorar las habilidades técnicas del personal entrenado. Sería deseable una segunda etapa de estas capacitaciones que incluya profusamente MLOps y la participación en proyectos reales que incluyan MLOps en instituciones del Estado.

RS: *¿En la práctica, ¿cuáles serían los beneficios de contar con tales personas expertas en MLOps?*

LMF: El país se encuentra en una carrera con otros países de la región para ser influyente en Latinoamérica en el desarrollo de software y en este ámbito, el Machine Learning y la ciencia de datos son elementos diferenciadores, por ello es importante tener un número importante de personas expertas en MLOps. 🌐

Sara Gallardo M. Periodista comunicadora, universidad Jorge Tadeo Lozano. Ha sido directora de las revistas Uno y Cero, Gestión empresarial y Acuc Noticias. Editora de Aló Computadores del diario El Tiempo. Redactora en las revistas Cambio 16, Cambio y Clase Empresarial. Coautora del libro "Lo que cuesta el abuso del poder". Ha sido corresponsal de la revista Infochannel de México; de los diarios La Prensa de Panamá y La Prensa Gráfica de El Salvador y corresponsal de la revista IN de Lanchile e investigadora en publicaciones culturales. Se ha desempeñado también como gerente de Comunicaciones y Servicio al Comensal en Inmaculada Guadalupe y amigos en Cía. S.A. (Andrés Carne de Res) y editora de Alfaomega Colombiana S.A.; en la actualidad es asesora y editora en escritura y producción de libros. Es editora de esta revista.

Encuesta Nacional “Machine Learning Operations y sus desafíos de implementación en Colombia”

El estudio sobre tecnologías inteligentes, realizado por ACIS a través de Internet, recibió respuesta de ochenta encuestados a nivel nacional.

DOI: 10.29236/sistemas.n165a4

Joseph A. Gallego M., PhD(c)

Fabio A. González O., PhD

La Asociación Colombiana de Ingenieros de Sistemas (Acis) adelantó una investigación a través de sus plataformas en Internet, con el propósito de indagar sobre el uso de tecnologías de la información enfocadas en Machine Learning Operations (MLOps) y explorar el impacto de su penetración en el ámbito empresarial colombiano.

Metodología

La encuesta se realizó a través de Internet, dirigida a profesionales

del sector TI a nivel nacional. Se basó en un formulario virtual de siete preguntas.

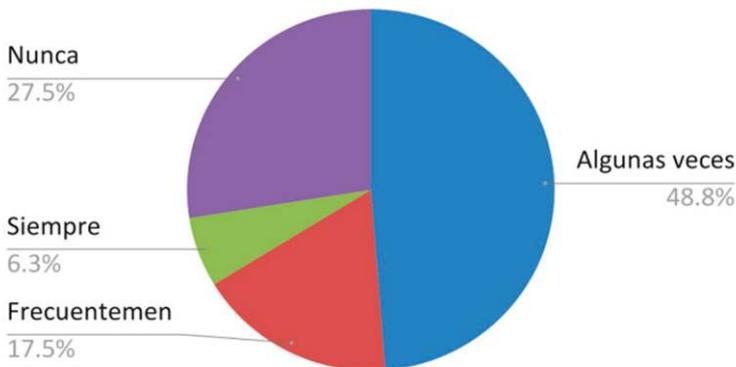
Tabla 1. Ficha técnica

Nombre de la investigación	Encuesta sobre Machine Learning Operations y sus desafíos de implementación en Colombia
Objetivo general	Identificar la percepción de los adecuados frente a las tecnologías inteligentes, su potencial de uso y la formación de los ingenieros para el uso de las mismas.
Realización	Asociación Colombiana de Ingenieros de Sistemas – ACIS
Tipo de investigación	Muestreo probabilístico
Población objetivo	Profesionales del sector de TI de Colombia.
Número de encuestados	80
Técnica de recolección	Encuesta online
Fecha de recolección	11 de Noviembre al 4 de Diciembre de 2022

Resultados

Se realizaron siete preguntas dirigidas a profesionales de TI y conforme a las respuestas se registraron los siguientes resultados:

1. ¿Su organización utiliza herramientas de análisis de datos y/o machine learning para apoyo a la toma de decisiones?



Gráfica 1. Grado de utilización de herramientas de análisis de datos y/o machine learning para apoyo a la toma de decisiones

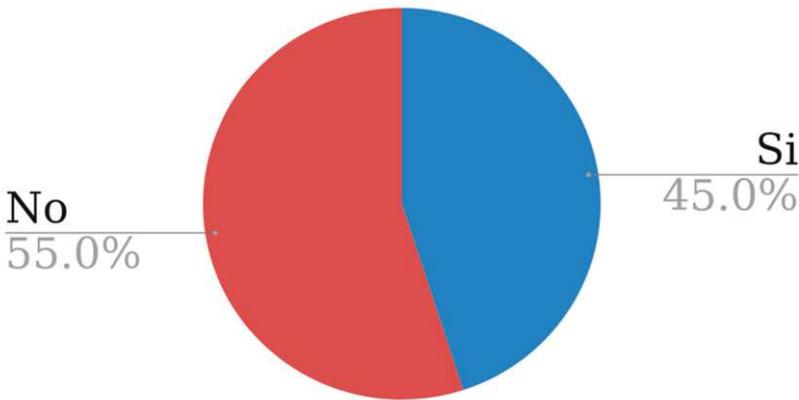
La mayoría de los encuestados, 48.8%, manifestó que algunas veces utiliza los datos y/o machine learning para apoyar su toma de decisiones. El 27.5%, indicó que nunca los usa. Sólo el 6.3% indicó que siempre los usa. De sus respuestas se puede inferir que las empresas colombianas aún tienen un gran potencial de mejoramiento basando sus decisiones estratégicas en datos.

2. ¿Su organización cuenta con un equipo especializado en el desarrollo de proyectos que involucran análisis de datos y/o machine learning?

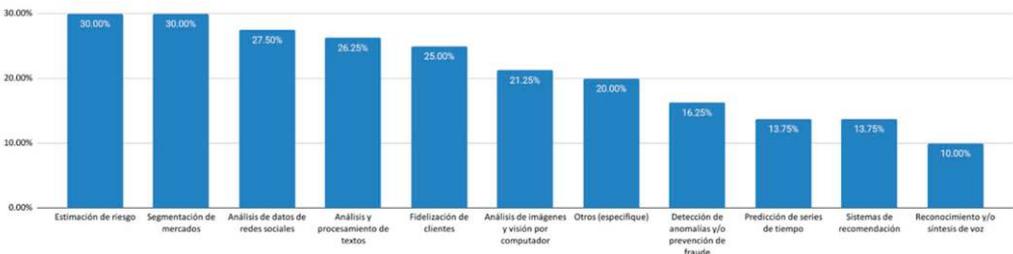
La gran mayoría de empresas, el 55%, no cuenta con un equipo especializado en proyectos de análisis de datos y/o machine learning. Tan sólo el 45% sí lo posee. Con esta respuesta, podemos ver que las empresas de tecnología encuestadas aún no ven la necesidad ni la rentabilidad que pueden generar los proyectos de análisis de datos y/o machine learning.

3. Cuáles de las siguientes aplicaciones de machine learning se han llevado a cabo en su organización.

Dentro de las aplicaciones de machine learning usadas por las orga-



Gráfica 2. Porcentaje de empresas con equipos de desarrollo de proyectos que involucran análisis de datos y/o machine learning



Gráfica 3. Uso de aplicaciones de machine learning dentro de la organización

nizaciones, de acuerdo con los encuestados, las de mayor uso son las relacionadas con el riesgo, la segmentación de mercados y el análisis de redes sociales. Técnicas de análisis de imágenes y visión por computador se usan en el 21% de las empresas encuestadas. Otras técnicas más avanzadas, como los sistemas de recomendación y el reconocimiento de voz, no son tan usadas.

4. ¿Su empresa utiliza Machine Learning Operations (MLOps)?

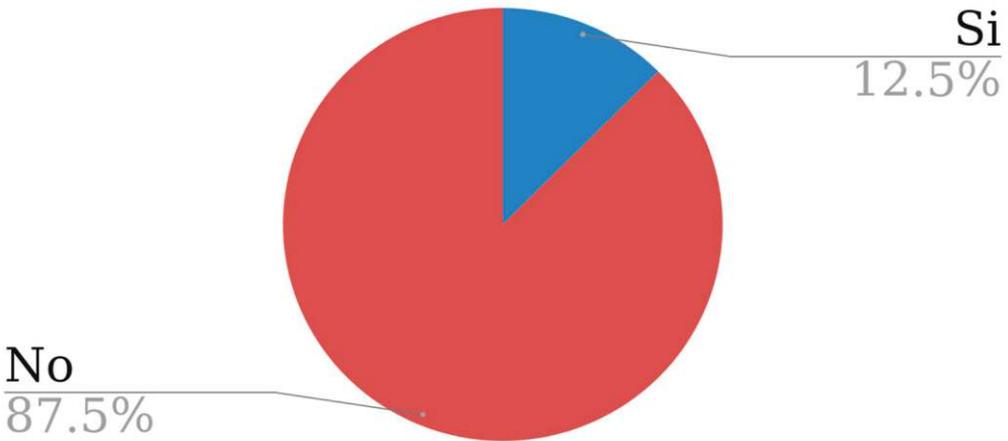
La gran mayoría de encuestados, 87.5%, no usa Machine Learning Operations (MLOps) dentro de sus empresas. Sólo el 12.5% lo utiliza. Podemos inferir que el uso de buenas prácticas de MLOps, todavía no se ve como una necesidad dentro de las empresas encuestadas de tecnología.

5. ¿Cuáles de los siguientes roles existen actualmente en su organización?

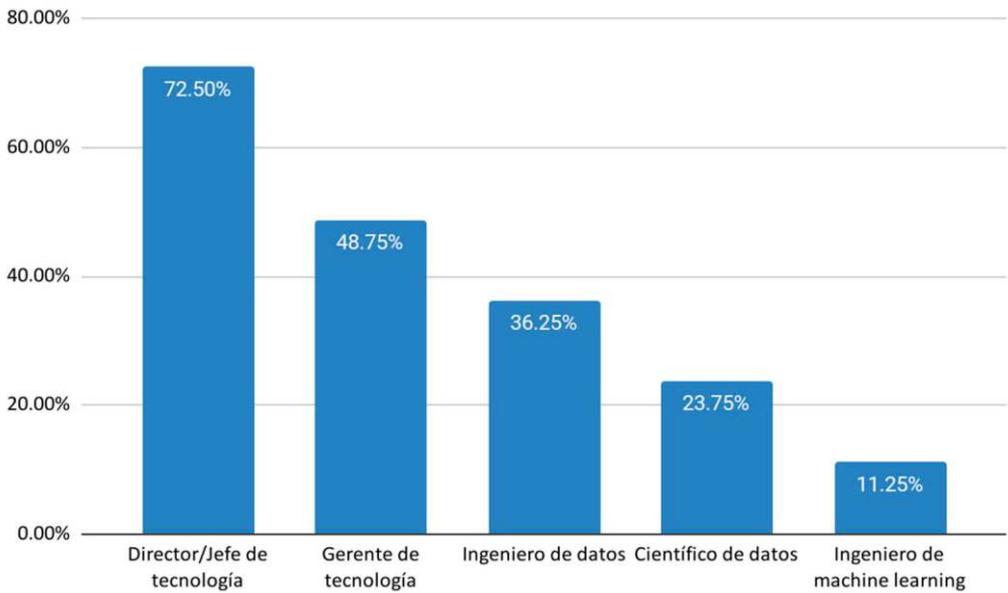
El 72.5% de las empresas encuestadas dice tener un director o jefe de tecnología. Sin embargo, sólo el 23.75% tiene un científico de datos, y el 11.25% un Ingeniero de machine learning. Las empresas encuestadas muestran que no se cuenta con el personal adecuado para desarrollar proyectos de ciencia de datos y machine learning.

6. ¿Se hacen uso de tecnologías de integración continua dentro de su organización?

El 51.3% de los encuestados usa tecnologías de integración continua frente al 48.8% que no hace uso de estas. El uso de tecnologías de integración facilita el desarrollo y mejora el tiempo de respuesta de



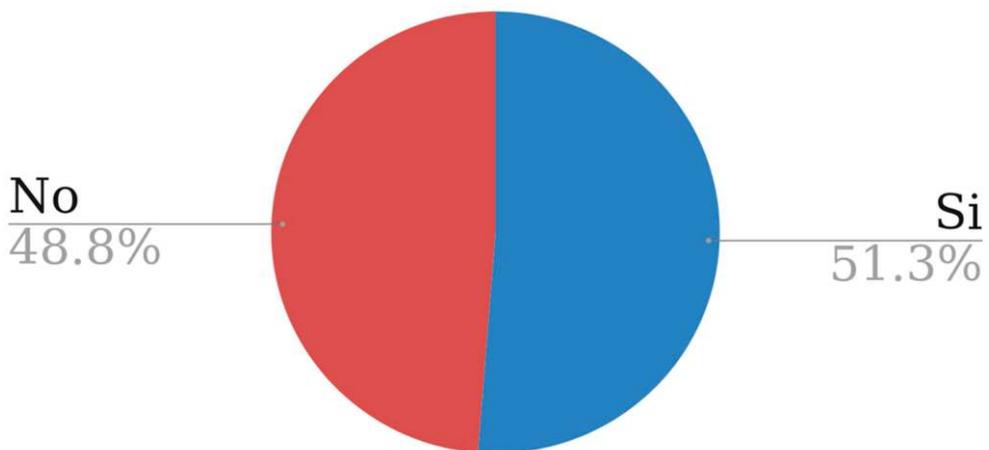
Gráfica 4. Uso de Machine Learning Operations (MLOps) dentro de empresas de tecnología en Colombia



Gráfica 5. Porcentaje por cada rol relacionado a MLOps dentro de las organizaciones encuestadas

las empresas. Existe todavía un gran número de empresas que no hace uso de este tipo de tecnologías.

7. ¿Qué herramientas tecnológicas para análisis de datos y/o machine learning se utilizan o se han utilizado en su organización?



Gráfica 6. Porcentaje de empresas encuestadas que hacen uso de tecnologías de integración continua.



Gráfica 7. Construida con nubedepalabras [1] según el uso de los encuestados para relacionar las tecnologías de análisis de datos y machine learning. A mayor tamaño de la palabra, mayor la frecuencia de la palabra utilizada los encuestados.

Con esta pregunta se buscaba indagar la penetración que tienen diferentes tecnologías relacionadas con análisis de datos y/o machine learning en el contexto colombiano. La siguiente nube de palabras muestra las respuestas más frecuentes.

Como se puede observar en la Gráfica 7, los encuestados consideran que conocen en mayor medida tecnologías de análisis de datos tales como PowerBi, Python, Microsoft Azure, Watson, Oracle, Sage Maker, SAS, entre otras.

Conclusiones

Los resultados muestran que existe un interés inicial dentro de las orga-

nizaciones encuestadas por utilizar el análisis de datos y machine learning dentro de sus organizaciones. No obstante, existe todavía una gran diferencia entre las empresas que utilizan los datos para tomar decisiones frente a los que no los usan o los usan algunas veces. El actual dinamismo de los negocios y los rápidos cambios en el consumo hacen que se deba actuar de forma rápida, esto se puede lograr utilizando métodos para toma de decisiones automatizadas como los de machine learning.

El uso de métodos de análisis de datos y machine learning es parcial dentro de las organizaciones. La mayoría de aplicaciones se enfo-

can en entender al cliente y en modelos financieros. Existe un gran potencial dentro de las organizaciones para utilizar métodos más avanzados de machine learning como la detección de anomalías, los sistemas de recomendación y el reconocimiento y/o síntesis de voz. Para lograr esto, las organizaciones deben empezar a utilizar herramientas automatizadas que permitan el rápido despliegue y pruebas automatizadas. Sin embargo, según los resultados de la encuesta cerca de la mitad de los encuestados no utiliza este tipo de herramientas. Así mismo, el uso de herramientas de MLOps que permitan el rápido despliegue de modelos de machine learning, su evaluación continua y monitorización es escaso dentro de los encuestados.

En el mundo se han ido creando nuevos roles dentro del análisis de

datos y el machine learning, que apuntan a tener un personal más especializado para este tipo de proyectos. Es el caso de los ingenieros de machine learning y los científicos de datos. Según se pudo constatar en la investigación, estos roles aún no son requeridos dentro de las organizaciones encuestadas. Esto deberá cambiar a futuro si queremos beneficiarnos del uso de métodos de aprendizaje automático.

En general, esta investigación mostró que existe un gran potencial para que las empresas empiecen a utilizar las herramientas de ML-Ops, con el propósito de crear modelos, probarlos, empaquetarlos, desplegarlos, y/o monitorizarlos.

Referencias

[1] NubeDePalabras:

<https://www.nubedepalabras.es/> 

Joseph Alejandro Gallego Mejía. Candidato a doctor en Ingeniería de Sistemas y Computación. Magíster en Ingeniería de Sistemas e Ingeniero de Sistemas y Computación de la Universidad Nacional de Colombia. Ingeniero Industrial de la Universidad Nacional de Colombia. Investigador en el área de las ciencias de la computación. Gerente de tecnología y fundador de Sammu, empresa internacional con presencia en Colombia, Chile, y Perú. Desarrollador web, móvil y back-end. Docente de la Universidad Nacional de Colombia, del programa Misión TIC 2022. Investigador asociado al grupo MindLab de la Universidad Nacional de Colombia, sede Bogotá; realiza investigación en aprendizaje de máquina, aprendizaje estadístico, minería de datos y análisis de datos.

Fabio Augusto González O. Profesor titular del Departamento de Ingeniería de Sistemas e Industrial de la Universidad Nacional de Colombia, sede Bogotá. Es ingeniero de Sistemas y Magíster en Matemáticas, de la Universidad Nacional de Colombia, MSc y PhD en Ciencias de la Computación de la Universidad de Memphis, Estados Unidos. Lidera el grupo de investigación MindLab. Su trabajo de investigación se enfoca en los fundamentos de aprendizaje de máquina y recuperación de información y su aplicación en el análisis de imágenes y texto, visión por computador y minería de datos.

Machine Learning Operations

DOI: 10.29236/sistemas.n165a5

Los desafíos de su implementación en Colombia.

Sara Gallardo M.

La producción de datos es cada vez mayor y su aumento seguirá en auge en computadores, teléfonos inteligentes, tabletas y la variedad de equipos que circulan entre los diferentes usuarios de la tecnología informática y de las comunicaciones, de ahí la importancia de profundizar en el tema.

Para tratar los asuntos más relevantes de este entorno en el país fueron convocados varios profesionales, quienes a través de su experiencia se ocuparon de analizar diferentes aspectos relacionados

con “Machine Learning” y su avance en el país.

El editor técnico y el coeditor de esta edición abrieron el debate.

Joseph Alejandro Gallego M.

Editor Técnico

Revista Sistemas Edición 165

Fabio Augusto González O.

Coeditor Técnico

Revista Sistemas Edición 165

En DevOps llamamos muro de la confusión al que se crea típicamente entre los desarrolladores de

software y los ingenieros de soporte del software ¿En su experiencia, en qué consiste el muro de la confusión en proyectos de machine learning y cómo MLOps ayuda a tumbar ese muro?

Raúl Ramos P.

*Profesor e investigador
Universidad de Antioquia*

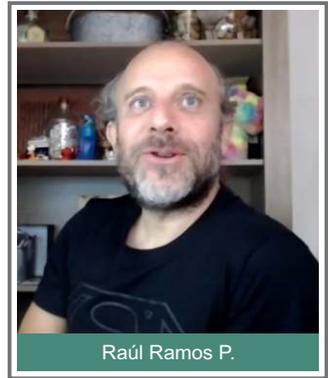
Con frecuencia hago una analogía que en esta ocasión me parece muy pertinente. Hoy todos estamos aprendiendo sobre cómo usar en la práctica y de manera útil todas estas técnicas de IA. El DevOps –sorprende que sea tan reciente- estaba en proceso de implementación

en ingeniería civil, mecánica, en una red de análisis y en prueba en los primeros proyectos de software. Hace cuarenta años había muchos problemas porque los requerimientos no eran del todo claros y se desarrollaron las metodologías ágiles. Con Machine Learning e IA está sucediendo lo mismo.

Ricardo Barragán O.

*Consultor Sénior
Teradata*

En el mundo de la analítica hay dos áreas que trabajan un poco en contravía; normalmente son áreas con personas innovadoras que están a la vanguardia, explorando



constantemente; no están basadas en procesos, sino en resultados. De ahí que un científico de datos no siga un proceso sino busca un resultado, sin importar qué herramientas tenga para lograrlo. Y, por el otro lado, están las áreas de tecnología, las cuales tienen que propender por la seguridad de una compañía, por la información, los ciclos, los tiempos y están basadas en el proceso. Se trata de dos áreas que trabajan en sentidos distintos y MLOps busca que trabajen juntas. Es decir, que exista la innovación, la flexibilidad, que la operación sea algo organizado y basado en procesos, es una forma de ver esas barreras.

Y otra forma que es importante mencionar, lo cual me recuerda un paper de Google que muestra cómo desarrollar un modelo analítico, es una parte pequeña con respecto a todo el proceso que se requiere para operacionalizar un modelo; es decir, no necesariamente cuando se termina de desarrollar un modelo, este es usable para la compañía.

Fabio A. González O.

Me parece muy interesante la perspectiva de Ricardo, en esa época ¿cómo se manejaba esa tensión, existía? Y ¿cómo se resolvía en proyectos de inteligencia de negocios, de ingeniería de datos?

Ricardo Barragán O.

Esa parte ha cambiado un poco; antiguamente, el 99% de los mo-

delos del área comercial, eran modelos descriptivos, principalmente era una visualización en una gráfica o en un reporte; hoy son otros y se trata de modelos predictivos o prescriptivos. Eso cambia un poquito la perspectiva y la operacionalización del modelo; se requiere tener en cuenta el aprendizaje y la ejecución del modelo. En mi compañía lo llamamos “scoring”, no se cómo lo denominen en la industria.

Fabio A. González O.



Es decir ¿eso tiene que ver con la complejidad de los modelos? ¿Esa brecha se ha agrandado?

Ricardo Barragán O.

Desde mi punto de vista sí hay más opciones, más herramientas para tener analítica y ahora las más usadas y las más difíciles de operacionalizar son las que contienen inteligencia artificial, el aprendizaje del modelo.

Diego Ibagón P.
Machine Learning Senior Manager
Mercado Libre



En compañías muy grandes de diferentes sectores el miedo inherente a llevar esta producción sigue existiendo y es curioso. Se explica también por la idiosincrasia. La relación academia empresa en los grupos empresariales grandes es muy lenta y nos quita la velocidad que sí tienen otros países, entonces ese miedo inherente genera el uso de enfoques antiguos, hace que uno trate de meter un triángulo en los juegos de los niños o un triángulo en la caja del cuadrado o el círculo en otra que no cuadra.

Jorge E. Camargo M.

El muro de la confusión en el ML-Ops es todavía mucho más complejo, es decir, estamos en una fase en que las empresas empiezan a adoptar esto. Si bien ya se está entrando a una adolescencia en el te-

ma del software, porque todavía no se puede decir adultez, se han organizado de alguna manera la casa y los roles; el tema todavía no se comprende muy bien, se especula mucho. De manera que un muro compuesto por ingeniería de software, más ciencia de datos, analítica de datos es pedregoso. Se trata de un tema en el que todavía estamos en pañales y en las entidades no se entiende muy bien.

Raúl Ramos P.



Me parece muy pertinente lo relacionado con los equipos multidisciplinarios y la parte del miedo, del pánico a la producción y creo que no es solamente una cuestión de cómo nivelar los equipos que están compuestos de muchas profesiones distintas; también es una cuestión cultural de cada profesión; no es igual, la parte del conocimiento, la manera de trabajar de un ingeniero de sistemas frente a la de un

físico o un economista. Cada vez que en un proyecto no está definido lo que hay que hacer, es necesario interrelacionar con el cliente y una variedad de profesiones. Es más viable en lugar de ingenieros de sistemas, optar por vincular físicos, en la medida en que los primeros requieren los asuntos muy bien definidos. No saben asumir el vocabulario de los otros ni manejar una terminología distinta a la propia. Esta cuestión de los equipos multidisciplinarios es muy difícil de poner en marcha.

Joseph Alejandro Gallego M.



En ingeniería de software se utiliza el término deuda técnica para referirnos a las consecuencias que conlleva el reprocesso, debido a soluciones rápidas a problemas complejos. Esto se debe a presiones internas y externas al equipo de trabajo.

En su experiencia como experto ¿cuáles son los principales riesgos

en los proyectos de Machine Learning que desencadenan en deuda técnica?

Ricardo Barragán O.



En mi opinión, veo los problemas de inteligencia artificial y Machine Learning un poco diferentes a los problemas de desarrollo que teníamos antes; de hecho, son más restantes. En todas las organizaciones se ve que el científico de datos está trabajando con una versión y el otro está trabajando con una diferente y el problema lo van a tener todos en el momento en el que quieran operacionalizar, a menos de que se haga independiente esta acción de los desarrollos. Debe haber una capa en el medio que independice eso.

Raúl Ramos P.

En mi opinión el concepto de deuda técnica aplicado a este tipo de entornos es un poco distinto también;

en la línea de lo que se está mencionando es bueno hacer ciclos rápidos, equivocarse pronto, gestionar las expectativas de los clientes, salir a producción lo antes posible, aunque sea parcialmente. El riesgo es no darse la oportunidad de hacer todo eso, porque en muchas ocasiones esto es también un ejercicio de descubrimiento, de validación de cuáles tecnologías y métodos pueden resolver el problema.

Fabio A. González O.

Muchos científicos de datos desarrollan buena parte de su trabajo usando herramientas interactivas como los cuadernos de Jupyter. ¿Cómo pueden estos profesionales evolucionar su trabajo hacia el uso de MLOps?, ¿en qué aspectos se deberían enfocar primero?

Jorge E. Camargo M.



Lo que uno encuentra con estos procesos tan nuevos en las com-

pañías es que los roles todavía no son claros, no están definidos y, muchas veces, las primeras personas que empiezan a abordar estos proyectos hacen de todo, es decir resultan aprendiendo machine learning, conectándose a la nube, subiendo máquinas. En otras palabras, haciendo de todo, porque no hay más profesionales en las compañías.

Raúl Ramos P.

En mi opinión, se trata de dos caminos muy distintos en términos de formación. Me parece que los profesionales deben acercarse para conocerse entre sí, es necesario interactuar para interpretar los datos y los modelos.

Ricardo Barragán O.

La falta de conocimiento profundo de las ramas requeridas siempre va a repercutir en costos. En una oportunidad trabajamos para un cliente con un científico de datos, una eminencia con amplios conocimientos, pero el consumo de los recursos era enorme y resultaba muy costoso para la compañía, de manera que se optó por un ingeniero que depuró el programa y quedó consumiendo el 80% menos de recursos. Es necesario el trabajo conjunto, cada uno con su aporte individual.

Diego Ibagón P.

Es importante generar conciencia en términos de costos y para ello es aconsejable trabajar con ejercicios muy didácticos de cara al diseño.

Fabio A. González O.

Sabemos que se trata de un tema nuevo del que todos estamos aprendiendo. ¿Qué se necesitaría para que una empresa pueda empezar a cambiar esa cultura? Desde las universidades a nivel de Ingeniería de Sistemas y otras áreas afines ¿qué nos faltaría?

Diego Ibagón P.

El asunto no es fácil, no se me ocurre la solución completa. Los ingenieros de sistemas adolecemos de muchas cosas. Los jóvenes, además de ver la solución en una forma más práctica, tienen conciencia del impacto y en ese sentido hay que volver a lo básico en muchos aspectos.

Raúl Ramos P.

Desde la perspectiva de la formación, en la Ingeniería de Sistemas es necesario involucrar la cultura de “poner cajitas juntas”, además de profundizar en diferentes aspectos. Así mismo, dar lugar a la prueba y el error. En las empresas es

importante entender que se trata de proyectos de largo aliento y que los resultados pueden demorar.

Ricardo Barragán O.

No estamos viendo los problemas a largo plazo sino simplemente el día a día en cuanto a la formación. En cuanto a las empresas deben innovar, aprender porque el mercado es quien valida el producto desarrollado.

Jorge E. Camargo M.

En todas las profesiones tenemos un déficit de formación importante, en muy pocos colegios se desarrolla el pensamiento computacional, asunto clave sin importar la carrera. En esa dirección, los retos son grandes.

Ricardo Barragán O.

Desde mi experiencia como proveedor, en primera instancia se debe analizar el problema y luego elegir la tecnología para solucionarlo. Por lo general, esto no sucede y el camino se recorre al revés. 🌐

Sara Gallardo M. Periodista comunicadora, universidad Jorge Tadeo Lozano. Ha sido directora de las revistas Uno y Cero, Gestión empresarial y Acuc Noticias. Editora de Aló Computadores del diario El Tiempo. Redactora en las revistas Cambio 16, Cambio y Clase Empresarial. Coautora del libro “Lo que cuesta el abuso del poder”. Ha sido corresponsal de la revista Infochannel de México; de los diarios La Prensa de Panamá y La Prensa Gráfica de El Salvador y corresponsal de la revista IN de Lanchile e investigadora en publicaciones culturales. Se ha desempeñado también como gerente de Comunicaciones y Servicio al Comensal en Inmaculada Guadalupe y amigos en Cía. S.A. (Andrés Carne de Res) y editora de Alfaomega Colombiana S.A.; en la actualidad es asesora en escritura y producción de libros. Es editora de esta revista.

Machine Learning

DOI: 10.29236/sistemas.n165a6

Como apoyo a la medicina predictiva.

Resumen

El sector salud tiene involucrado una gran cantidad de procesos y procedimientos generadores de todo tipo de información que en muchos casos no están disponibles de forma libre para los profesionales de diferentes áreas y en especial de las ciencias computacionales.

¿Qué sucedería si toda esta información pudiera estar disponible? La medicina preventiva y predictiva podría desarrollarse con mayor rapidez, desarrollando modelos predictivos a través de algoritmos de Machine Learning, como apoyo a los profesionales de la salud en la toma de decisiones.

Este artículo permite conocer la convergencia que existe entre la medicina predictiva y el Machine Learning, sus ventajas y los diferentes algoritmos de Machine Learning que se pueden aplicar dependiendo de los tipos de datos.

Palabras claves

Machine Learning, Health, Inteligencia artificial

Introducción

La Organización Mundial de La Salud (OMS) ha planteado desde 19-97 la integración de las TIC para superar la barrera de acceso y aumentar la cobertura de los servicios de salud. El término e-Salud o e-Health se ha definido como el “soporte que las Tecnologías de la Información y las Comunicaciones, ofrecen a la salud y a los ámbitos relacionados con ella, buscando tanto la inclusión de los servicios de atención de salud, la vigilancia y la documentación sanitarias, como la educación, los conocimientos y las investigaciones en materia de salud”, según lo contemplado en el documento “Estrategias y Plan de Acción sobre e-Salud”, publicado en el 2011 por la OMS. De acuerdo con la OMS, la e-Salud se divide en varios componentes:

- 1) las historias clínicas electrónicas, definidas como un formato electrónico de información sobre la salud del paciente;
- 2) la telemedicina, que consiste en la prestación de servicios de salud utilizando las TIC;
- 3) la m-Salud o el acceso a la prestación de servicios en salud mediante dispositivos móviles, que es un término empleado para designar el ejercicio de la medicina y la salud pública con el apoyo de dispositi-

vos móviles, según (Villamizar & Lobo, 2019). Por medio de la integración de los servicios de e-Salud con la inteligencia Artificial (IA), se podrán detectar de manera precoz diferentes tipos de enfermedades, como apoyo a los profesionales de la salud y las entidades de gubernamentales, en la toma de decisiones en el momento de una crisis de salud pública como los ocurridos con la pandemia de Covid-19.

La IA según (Romero Cardalda, Da-fonte Vázquez, & Gómez García, 2007), el aprendizaje, la capacidad de adaptación a entornos cambiantes y la creatividad, entre otros aspectos, son facetas que usualmente se relacionan con el comportamiento inteligente. La Inteligencia Artificial (IA) es por definición interdisciplinar, porque en ella intervienen diferentes disciplinas, tan variadas como la neurociencia, la psicología, las tecnologías de la información, la ciencia cognitiva, la física, las Matemáticas y la medicina. Particularmente, en el caso de la medicina la IA, constituye uno de los campos interdisciplinarios y transfronterizos donde convergen muchas ciencias. La aparición de las computadoras y la elaboración de las teorías de la computación, la información y el control, proporcionaron los soportes experimentales y teóricos para la investigación en el área de la IA Figura 1. Muchas de

las esferas que emplean estos equipos, incluidas las ciencias médicas, la consideran esencial entre sus líneas estratégicas de investigación y entre las que se convierten en factor de progreso, porque como algunos autores expresan deben gran parte de su actual desarrollo a los resultados obtenidos en el proceso de cierto tipo de problemas médicos: el diagnóstico y el tratamiento de diversas enfermedades (Honavar, 2006). Antes de hablar como convergen el Machine Learning con la medicina predictiva, definiremos algunos términos esenciales de IA.

Machine Learning (ML) o aprendizaje automático es un subcampo de la IA que incluye algoritmos de autoaprendizajes que derivan el conocimiento a partir de los datos

para crear predicciones. El ML ofrece alternativas eficientes para la captura de conocimientos en datos, mejorar los modelos predictivos y tomar decisiones con base en esos datos. En el ML existen tres tipos Figura 2, el aprendizaje supervisado, el aprendizaje no supervisado y el aprendizaje reforzado. (Dorado-Díaz, 2019)

Aprendizaje supervisado

Su principal objetivo y el eje central es el aprendizaje de un modelo, a partir de una serie de datos etiquetados, el cual permite hacer predicciones a partir de una serie de datos futuros. Es muy importantes que los datos deben estar ordenados, etiquetados y limpios; es decir, se debe hacer un proceso de ETL (extracción, transformación y carga). El aprendizaje de supervisa-



Figura No 1. Inteligencia artificial. Fuente: Autores.

Deep Learning:

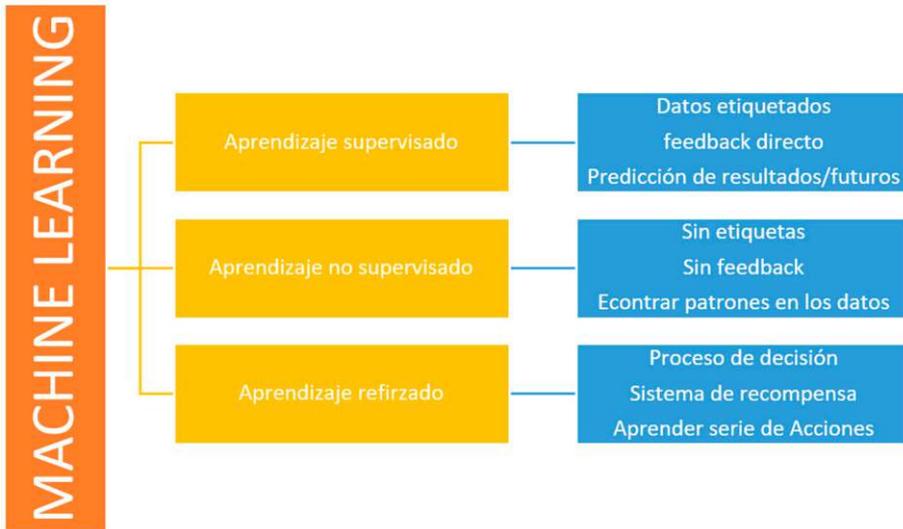


Figura No 2: Tipos de machine Learning. Fuente: autores

do se utiliza en muchos casos de la medicina, por ejemplo, se puede utilizar un modelo para relacionar características de lesiones de piel utilizando variables como ABCDE (Asimetría, Bordes, Color, Diámetro y Evolución) clasificando la lesión como maligna o benigna.

Aprendizaje no supervisado

A diferencia del aprendizaje supervisado, el no supervisado no proporciona datos previos, más bien se carga una gran cantidad de datos no etiquetados y el modelo se encarga de encontrar patrones similares ocultos, el cual separa los datos en grupos de manera automática, identificar datos atípicos o producir representaciones de baja dimensión de datos, es decir, reducción del número de variables aleatorias. En el campo de la medi-

cina se puede aplicar en el procesamiento de imágenes para separar tumores de acuerdo a los patrones encontrados.

Aprendizaje reforzado

Su principal característica es el aprendizaje desde su propia experiencia; es decir, tomar la mejor decisión ante diferentes situaciones de acuerdo con el proceso de prueba y error, donde se recompensan las decisiones correctas. Este tipo de aprendizaje se está utilizando en el reconocimiento facial, diagnósticos médicos y clasificaciones de ADN.

¿Cómo se escoge el algoritmo correcto de Machine Learning para resolver un problema?

Para seleccionar el algoritmo más adecuado para una solución de

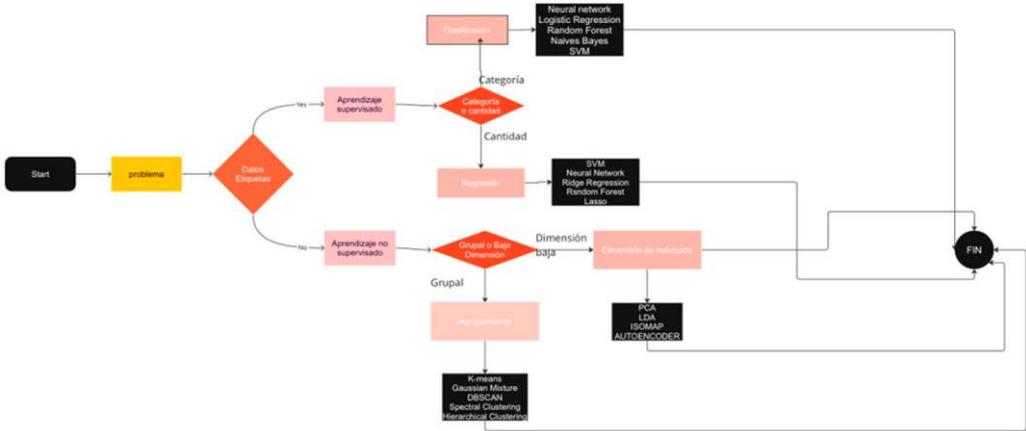


Figura No 3: Selección el algoritmo correcto de Machine Learning.
Fuente: Autores

Machine Learning se debe hacer lo que se indica en la figura No 3.

Se debe observar si los datos están etiquetados o no, de acuerdo con eso se selecciona si es aprendizaje supervisado o no supervisado. Si es aprendizaje supervisado, se tiene en cuenta si es clasificación (categoría) o regresión (cantidad), y se selecciona el o los algoritmos para desarrollar el modelo. Si se escoge aprendizaje no supervisado, se debe ver si es de agrupamiento o dimensión y se seleccionan el o los algoritmos para el desarrollo del modelo.

Machine Learning como apoyo en la medicina predictiva

El mundo de hoy está colmado de una “sobredosis” de información, en donde el poder de los datos es considerado el nuevo petróleo; anteriormente el dato era utilizado

únicamente para actividades donde se representará información. Hoy en día esta gran cantidad de datos son utilizados en diferentes áreas del conocimiento como son el sector financiero, gobierno, marketing y ventas, transporte y área de la salud para predecir hechos. Según el Artificial Intelligence Institute de la Universidad de Stanford, existe un interés importante en la contratación de profesionales con conocimiento y manejo del Machine Learning en los Estados Unidos, donde la contratación aumento del 0,07% en el 2010 a más de un 0.51% en el 2019. ver figura No 4. El área de la salud es uno de los campos en donde más se ha aplicado el Machine Learning, en especial en la medicina predictiva, el cual se puede definir como el campo de la medicina que consiste en predecir o intuir medidas preventivas de enfermedades del paciente;

respondiendo a preguntas como ¿Qué sucedió? (Diagnostica), ¿Qué pasará (predecir) y ¿qué deberá hacer? (prescribir). El Machine Learning converge con gran éxito con la medicina predictiva, y se han logrado grandes avances en el diagnóstico precoz de enfermedades de los pacientes.

La sanidad y el área de la salud en general, es un sector que constantemente avanza e incorpora los avances tecnológicos más recientes en forma de nuevos dispositivos, nuevos descubrimientos científicos y su aplicación a la prevención, diagnóstico y tratamiento de dolencias y enfermedades. En años recientes el Machine Learning se ha incorporado a esta sinergia

entre tecnología y ciencia del sector de la salud, ayudando y potenciando a los profesionales médicos en su labor.

En el área de la salud, el Machine Learning es una valiosa herramienta para el análisis de datos y ayuda al diagnóstico, mediante el uso de sus algoritmos de clasificación y predicción, ayudando a los profesionales de la salud, en responder más rápidamente a los posibles eventos médicos de los pacientes y detectar patrones epidemiológicos, que si se analizaran los datos de forma manual.

Un reto importante es lograr que los profesionales médicos y administrativos del sector de la salud apro-

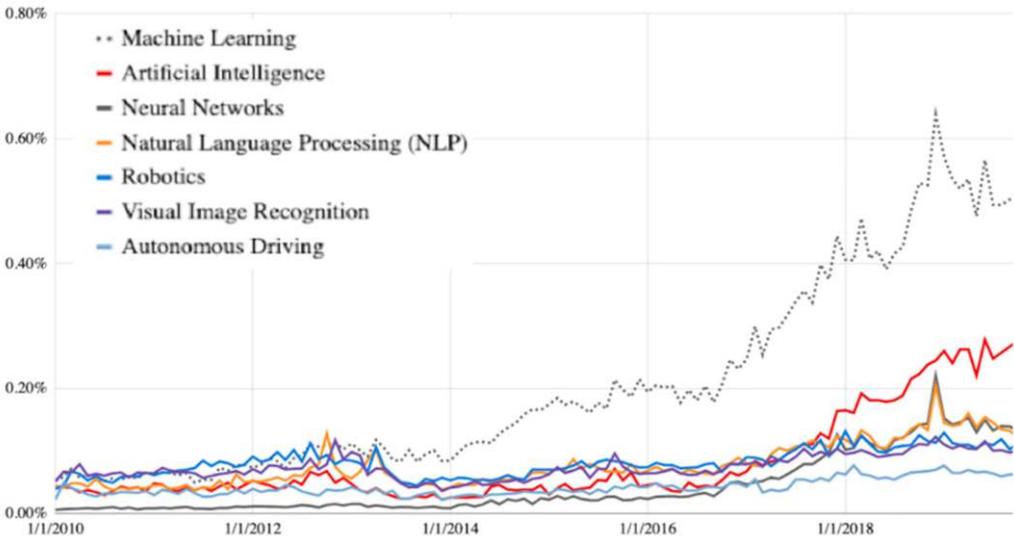


Figura No 4. Crecimiento de las ofertas de empleo relacionadas con inteligencia artificial en Estados Unidos desde 2010. Fuente: <https://www.revistacuantica.com/index.php/rcq/article/view/27/29>

pien más rápidamente las cada vez más comunes y efectivas aplicaciones del Machine Learning, herramientas que permiten agilizar la respuesta del sistema de salud a los pacientes, disminuir la carga del análisis especializado de los médicos, automatizar procesos administrativos y todo ello apuntando como meta principal a mejorar la experiencia del paciente en su atención y optimizar los recursos en la prestación del servicio por parte de los operadores del sistema de salud.

Uno de los campos en donde encontramos desarrollo de aplicaciones utilizando Machine Learning es el de ayudas diagnósticas empleando imágenes médicas, en el cual encontramos múltiples casos, desde aplicaciones en etapa de desarrollo en centros de investigación y universidades, hasta algunas ya en uso por parte de algunos centros médicos. El proyecto de análisis automatizado de imágenes de la retina para la detección de la retinopatía diabética referible fue el pionero en este campo, este análisis automatizado de imágenes de la retina tiene una alta sensibilidad y especificidad para detectar la retinopatía diabética referible y pudo implementarse de forma segura en el proceso de tamizaje de la retinopatía diabética, mejorando el acceso al tamizaje y reducir la pérdida visual mediante un tratamiento temprano. (Abràmoff, Folk., & Han, 2013). Podemos mencionar algunos de estos casos, como la detec-

ción de enfermedades de los ojos, utilizando el Deep Learning a partir de imágenes oftalmológicas, en este tenemos por ejemplo el proyecto liderado por Lily Peng de Google (DSW Ting, 2019).

También podemos mencionar los proyectos y las aplicaciones desarrolladas por la universidad de Stanford para la detección de Cáncer de piel mediante fotografías de la lesión (Esteva, y otros, 2017) (Trinh, y otros, 2022), la aplicación UMSkinCheck de la Universidad de Michigan, que a partir del uso de Machine Learning permite realizar el auto examen de cáncer de piel y hacer un seguimiento al almacenar las fotografías de las lesiones de los usuarios, además incluye una calculadora de riesgo de melanoma. Las personas que están en un alto riesgo de cáncer de piel se les indica que deben realizarse autoexámenes frecuentes de la piel y a la vez, realizar fotografías de todo su cuerpo, para llevar un historial de lunares y lesiones de la piel. University of Michigan Health, 2022).

Para el diagnóstico de enfermedades cancerígenas, también existe en el mercado Skin visión, la cual es una aplicación móvil que combina la tecnología de IA, Machine Learning y procesamiento de imágenes con la experiencia de profesionales de área de la salud de la piel; la aplicación consiste en la detección precisa y oportuna del cáncer de piel, junto con asesoramiento personalizado al paciente y reco-

mendación de rutas de salud. La aplicación Skin Visión tiene un 95% de precisión en la detección de cáncer de piel. (Udrea, y otros, 2019). Existen muchas otras investigaciones en la misma área de la salud como precisión de una aplicación de teléfono inteligente para clasificar lesiones cutáneas basada en algoritmos de aprendizaje automático, entre ellas podemos mencionar la de Segmentación automatizada de lesiones cutáneas en imágenes dermatoscópicas mediante los algoritmos GrabCut y k-means (Murugaiyan Jaisakthi, Mirun, & Aravindan, 2018) y Aprendizaje profundo basado en wavelets para la clasificación de lesiones cutáneas (Sarte & Hasan, 2020)

En el Hospital Infantil de Cincinnati desarrollaron un proyecto de investigación de Machine Learning aplicado al suicidio, a partir de un algoritmo de clasificación y regresión en donde se categorizaron tres grupos así: personas con riesgo de suicidio, personas con trastornos mentales sin riesgo de suicidio y un grupo sin riesgo de suicidio ni trastornos mentales, con 93% de precisión para indicar si una persona posee riesgo de suicidio o no. (Cruz Micán, Poveda Aguja, & Buitrago Márquez, 2020).

Con el auge en el uso de las redes sociales, encontramos diversos proyectos de investigación que trabajan en la identificación de personas con riesgo de suicidio a partir del análisis de sus publicaciones en

redes sociales. (Ji, Ping Yu, Sai-fu, Pan, & Guodong, 2018) (Sawhney, Manchanda, Mathur, & Singh, 2018), las compañías que gestionan las redes sociales como Facebook, también se encuentran trabajando en el desarrollo de algoritmos para la detección de personas con riesgo de suicidio (Palazuelos, 2017) (Meta, 2022).

En Colombia existen proyectos en los que se está trabajando en la identificación de personas con riesgo de suicidio como el proyecto de investigación de un Algoritmo para la detección del suicidio mediante el análisis de datos en las redes. (Saldarriaga Gómez, 2021)

A nivel nacional se desarrolló un proyecto llamado transformación digital del sistema de salud colombiano y el Ministerio de Salud busca adaptarlo a un mejor esquema de gestión apoyado en varias tendencias tecnológicas actuales. El proyecto denominado Agenta TD-Salud, busca integrar varias tecnologías como historias clínicas electrónicas, telesalud, análisis de Big Data, (Cruz Micán, Poveda Aguja, & Buitrago Márquez, 2020)

En general, el Machine Learning puede utilizarse para desarrollar mejores herramientas de diagnóstico utilizando imágenes médicas y resultados de pruebas de laboratorio, que permitan de forma más rápida y certera, obtener un diagnóstico para el paciente. Por ejemplo, un algoritmo de Machine Lear-

ning a partir de imágenes médicas, que pueden incluir fotografías especializadas, imágenes de ultrasonido, imágenes radiológicas o resonancias magnéticas, y mediante el uso de sus técnicas para el reconocimiento de patrones y clasificación, obtener características que indiquen la presencia de una anomalía o de una enfermedad en el paciente.

Los enfoques de Machine Learning pueden ser utilizados para modelar datos epidemiológicos, mejorando nuestra comprensión de la distribución, frecuencia y factores determinantes de las enfermedades existentes en núcleos de población, encontrando patrones de dispersión y avance de enfermedades y realizando predicciones de evolución.

El Machine Learning también acelera el desarrollo de nuevos tratamientos y puede utilizarse para identificar información relevante de toxicidad, inocuidad, combinaciones, entre otros, en los datos ya existentes que podrían conducir al desarrollo de nuevos medicamentos y tratamientos para las enfermedades.

Reflexiones finales

Aunque los algoritmos de Machine Learning existen desde hace mucho tiempo, en los últimos años han tenido mayor uso al volverse más populares, debido al desarrollo de herramientas que facilitan su uso y a la aparición de novedosas técni-

cas, fundamentalmente en el Deep Learning. En el área de la salud hemos podido encontrar numerosas aplicaciones que van más allá de las aplicaciones académicas y algunas de ellas ya se encuentran en uso frecuente por los profesionales de la salud y otras que han sido puestas al servicio de las personas para un autodiagnóstico, aunque se debe tener precaución con estas últimas por los potenciales riesgos que pueden generar.

Al utilizar el Machine Learning y su combinación con otras tecnologías médicas de pruebas y exámenes diagnósticos, es posible diagnosticar enfermedades más rápidamente y a un menor costo que las técnicas de evaluación manual actuales; además, potenciará el uso de estrategias de telemedicina, como opción a la visita presencial. Así mismo permitirá desarrollar programas de tamizaje a un menor costo, atendiendo grupos de población cada vez más amplios.

Entre los desafíos clave en el desarrollo e implementación de proyectos con Machine Learning está la recopilación de datos de calidad. Los temas normativos también presentan desafíos importantes para lograr una amplia difusión y uso de estas herramientas. En Colombia aún no tenemos una legislación clara y completa en este campo. Así mismo es importante que los profesionales de la salud, puedan ver estas tecnologías como herramientas de ayuda y soporte en su

labor y no como un elemento de remplazo del hombre por la máquina.

Los principales proveedores de tecnología empresarial en la nube también aportan herramientas para desarrollar soluciones que involucran Machine Learning, desde la recopilación de los datos hasta el despliegue y uso de aplicaciones.

Referencias

- Abràmoff, M., Folk, J., & Han, D. (2013). Automated Analysis of Retinal Images for Detection of Referable Diabetic Retinopathy. *JAMA Ophthalmology*.
- Cruz Micán, E., Poveda Aguja, F., & Buitrago Márquez, L. (2020). Las TIC en el sector salud, machine learning para el diagnóstico y prevención de enfermedades. *Revista científica Quantica*, 1-32.
- Dorado-Díaz, P. I.-G.-P. (2019). Aplicaciones de la inteligencia artificial en cardiología: el futuro ya está aquí. *Revista Española de Cardiología*, 72(12), 1065-1075.
- DSW Ting, L. P. (2019). Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology*, 167-175.
- Esteva, A., Kuprel, B., Novoa, R., Ko, J., Swetter, S., Blau, H., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 115-118.
- Honavar, V. (2006). *Artificial Intelligence An Overview*. Ames: Iowa State University.
- Ji, S., Ping Yu, C., Sai-fu, F., Pan, S., & Guodong, L. (2018). Supervised Learning for Suicidal Ideation Detection in Online User Content. *Social Big Data: Mining, Applications, and Beyond*.
- Meta. (30 de 11 de 2022). Under the hood: Suicide prevention tools powered by AI. Obtenido de <https://ai.facebook.com/blog/under-the-hood-suicide-prevention-tools-powered-by-ai/>
- Murugaiyan Jaisakthi, S., Mirun, P., & Aravindan, C. (2018). Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms. *IET Computer Vision*, 1088-1095.
- Palazuelos, F. (27 de 11 de 2017). Facebook desarrolla un algoritmo para detectar conductas suicidas antes que los humanos. *El país*.
- Romero Cardalda, J., Dafonte Vázquez, C., & Gómez García, A. (2007). *Inteligencia Artificial y Computación Avanzada*. Santiago de Compostela: Fundación Alfredo Brañas.
- Saldarriaga Gómez, P. (2021). Algoritmo para la detección del suicidio mediante el análisis de datos en las redes. Tesis de grado. Medellín: Tecnológico de Antioquia.
- Sarte, S., & Hasan, D. (2020). Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms. *IET Computer Vision*, 720-726.
- Sawhney, R., Manchanda, P., Mathur, P., & Singh, R. (2018). Exploring and Learning Suicidal Ideation Connotations on Social Media with Deep Learning. *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 167-175.
- Trinh, P., Yekrang, K., Phung, M., Pugliese, S., Chang, A., Bailey, E., . . . Sarin, K. (2022). Partnering with a senior living community to optimise teledermatology via full body skin screening during the COVID-19 pandemic: A pilot programme. *Skin Health and disease*.
- Udrea, A., Mitra, G., Costea, D., Noels, E., Wakkee, M., de Carvalho, T., & Nijsten, T.

(2019). Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. *Journal of the European Academy of Dermatology and Venereology*, 648-655.

University of Michigan Health. (2022). Skin Cancer Self-Exam Mobile App. Obtenido de

<https://www.uofmhealth.org/patient%20and%20visitor%20guide/my-skin-check-app>

Villamizar, A., & Lobo, R. (2019). Antecedentes y experiencias de e-Salud en Colombia. *REVISTA COLOMBIANA DE COMPUTACIÓN*, Volumen 17, numero 2, 76–89. 

Luis Alfredo Blanquicett Benavides: *Nació en Cartagena Colombia en 1979. Recibió el grado de Ingeniero de Sistemas de la Universidad Rafael Núñez en el 2005, especialista en Gerencia de Sistemas de Información de la Universidad del Norte en el 2007 y grado de Máster en Dirección Estratégica de Sistemas de Información en el 2015 de la Universidad Internacional de Puerto Rico, candidato a doctor en informática en la universidad de Oviedo -España. Se ha desempeñado en el desarrollo de proyectos de investigación de inteligencia artificial y analítica de datos con la corporación universitaria Rafael Núñez y la universidad del Sinú y desarrollo de proyectos en consultoría externa en pruebas de software.*

Luis Fernando Murillo Fernández: *Ingeniero Electricista de la Universidad Tecnológica de Bolívar, especialista en Informatica Industrial de la Universidad del Cauca y Máster en Gestión de la Innovación de la Universidad Tecnológica de Bolívar, candidato a doctor en Energía y control de procesos en la universidad de Oviedo -España. Se ha desempeñado en el desarrollo de proyectos de investigación de inteligencia artificial y analítica de datos, y en sistemas de control e instrumentación.*

Queremos invitarte a
que te vincules como
PATROCINADOR de
nuestros eventos

APOYANOS

PARA MAS INFORMACIÓN
EN :
WWW.ACIS.ORG.CO
3015530540 - 3043463413

O
ESCRIBENOS A:
CURSOS@ACIS.ORG.CO

