



UNIVERSIDAD ESTATAL DE BOLÍVAR

**FACULTAD DE CIENCIAS ADMINISTRATIVAS, GESTIÓN EMPRESARIAL E
INFORMÁTICA**

CARRERA DE SOFTWARE

**TRABAJO DE INTEGRACIÓN CURRICULAR
PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN SOFTWARE**

FORMA: PROYECTO DE INVESTIGACIÓN

TEMA:

EL APRENDIZAJE AUTOMÁTICO APLICADO AL DESARROLLO DE SOFTWARE.

AUTOR:

RICARDO JAVIER CHORA PAUCAR

DIRECTOR:

ING. EDGAR RIVADENEIRA RAMOS

GUARANDA – ECUADOR

2024

TEMA DEL PROYECTO DE INVESTIGACIÓN

EL APRENDIZAJE AUTOMÁTICO APLICADO AL DESARROLLO DE SOFTWARE.

AGRADECIMIENTO

Quiero agradecer a Dios por bendecirme y nunca dejarme solo en cada paso de mi vida, sobre todo en la etapa universitaria, el me iluminó para darme esa claridad y enfrentar cada desafío que se me presentaba en el camino, me bendijo con perseverancia para nunca rendirme y seguir con pie firme para poder llegar a la meta final que es la de graduarme.

A mis padres, por ser esos pilares de cada uno de mis logros. Gracias por enseñarme con su ejemplo, que con la perseverancia y mucho esfuerzo uno puede alcanzar cualquier meta en la vida; por estar en cada paso, guiándome y acompañándome con su apoyo incondicional.

Gracias papá y mamá por siempre confiar en mí, este logro es tan suyo como mío.

A mi tutor y mis pares, por su valiosa orientación, paciencia y compromiso durante el desarrollo de esta tesis. Sus consejos y sus experiencias han sido fundamentales para el logro de este trabajo.

A todos mis profesores, quienes a lo largo de mi formación compartieron su conocimiento y pasión por el aprendizaje, dejándome valiosas enseñanzas que me acompañarán en mi vida profesional.

Finalmente, agradezco a mis amigos y compañeros de universidad, quienes, con su compañía, palabras de ánimo y momentos compartidos hicieron de este viaje una experiencia inolvidable.

DEDICATORIA

Dedico esta tesis, primero a Dios, quien me ha dado la fuerza, la paciencia y la sabiduría para aprender de cada experiencia, este logro te lo dedico a ti con gratitud y humildad, sabiendo que, sin tu apoyo, nada de esto habría sido posible.

A mis padres, por ser mi mayor inspiración por su amor incondicional, su apoyo inagotable y por creer en mí incluso en los momentos más difíciles; a la vez pido perdón por esas muchas veces que me solía perder en la Universidad solo por pensar que todo era diversión y fiestas, creyendo que todo lo podía solo, sin pensar en cómo ustedes, en silencio, siempre estaban ahí, preocupados, esperando una llamada, un mensaje, una señal de que todo está bien. Perdón por las veces que me equivoqué y no regresé a ustedes con humildad, por pensar que el tiempo todo lo solucionaba, sin entender que un simple “lo siento” habría aliviado sus corazones. Gracias por ese amor tan incondicional, incluso cuando no fui el hijo más recíproco. Su paciencia y comprensión son mi mayor lección, y con esta tesis, espero honrar al menos una pequeña parte de lo que siempre han hecho por mí. Gracias papá y mamá por todo, a ustedes les debo todo lo que soy y todo lo que seré.

CERTIFICADO DE VALIDACIÓN**DERECHOS DE AUTOR** (Acorde al modelo emitido por biblioteca general)

CERTIFICADO DE VALIDACIÓN



FACULTAD DE CIENCIAS
ADMINISTRATIVAS,
GESTIÓN EMPRESARIAL
E INFORMÁTICA

CERTIFICADO DE VALIDACIÓN

ING. EDGAR RIVADENEIRA, ING. RODRIGO DEL POZO y FIS. RAFAEL MEDINA, en su orden de Director y Pares Académicos del Trabajo de Integración Curricular “EL APRENDIZAJE AUTOMÁTICO APLICADO AL DESARROLLO DE SOFTWARE.”, desarrollado por el estudiante RICARDO JAVIER CHORA PAUCAR.

CERTIFICAN

Que, luego de revisado el Trabajo de Integración Curricular en su totalidad, cumple con las exigencias académicas de la carrera SOFTWARE, por lo tanto, autorizamos su presentación y defensa.

Guaranda, 9 del 12 del 2024



ING.EDGAR RIVADENEIRA

Director



ING.RODRIGO DEL POZO

Par Académico



FIS.RAFAEL MEDINA

Par Académico

DERECHOS DE AUTOR

Yo RICARDO JAVIER CHORA PAUCAR portador de la Cédula de Identidad No 0930579594 en calidad de autor y titular de los derechos morales y patrimoniales del Trabajo de Titulación: EL APRENDIZAJE AUTOMÁTICO APLICADO AL DESARROLLO DE SOFTWARE, modalidad Proyecto Investigativo, de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN, concedemos a favor de la Universidad Estatal de Bolívar, una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos. Conservamos a mi favor todos los derechos de autor sobre la obra, establecidos en la normativa citada.

Así mismo, autorizo a la Universidad Estatal de Bolívar, para que realice la digitalización y publicación de este trabajo de titulación en el Repositorio Digital, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

El autor declara que la obra objeto de la presente autorización es original en su forma de expresión y no infringe el derecho de autor de terceros, asumiendo la responsabilidad por cualquier reclamación que pudiera presentarse por esta causa y liberando a la Universidad de toda responsabilidad.

Ricardo Chora

Ricardo Javier Chora Paucar
C.I. 0930579594

ÍNDICE DE CONTENIDO

INTRODUCCIÓN	1
RESUMEN	4
ABSTRACT	5
CAPÍTULO I.....	6
FORMULACIÓN GENERAL DEL PROYECTO	6
Descripción del Problema	6
Formulación del Problema	7
Preguntas de Investigación	7
Justificación	8
Línea de Investigación.....	9
Objetivos: General y Específicos.....	9
Idea a Defender	10
Variables	10
CAPÍTULO II MARCO TEÓRICO	11
Antecedentes	11
Marco Científico	13
Aprendizaje Automático (AA).....	13
Evolución del Aprendizaje Automático.....	13
Técnicas de Aprendizaje Automático.....	15
Desarrollo de software.....	17
Evolución del desarrollo de software.....	17
Metodología Prisma.....	18

Marco Conceptual	18
Aprendizaje Automático o Machine Learning (AA/ML):	18
Desarrollo de Software:	18
Marco Legal	19
CAPITULO III METODOLOGÍA	20
Tipo de Investigación	20
Investigación descriptiva	20
Enfoque de la investigación	20
Métodos de Investigación	21
Método inductivo	21
Técnicas e Instrumentos de Recopilación de Datos	21
Bibliográfica/Documental	21
Procesamiento de la Información	21
CAPITULO IV RESULTADOS Y DISCUSIÓN	22
Análisis, Interpretación y Discusión de Resultados	22
Metodología PRISMA.	22
Criterios de inclusión y de exclusión	22
2. Estrategia de Búsqueda:	22
Registro de la Búsqueda:	23
3. Selección de Estudios:	24
Análisis y Síntesis de los Resultados.	26
¿Cuál es el nivel de adopción del aprendizaje automático para el desarrollo de software?	26

¿De qué manera se ha aplicado el aprendizaje automático en el proceso de desarrollo de software?	28
¿Cuáles son las limitaciones para la adopción del aprendizaje automático en el proceso de desarrollo de software?	30
Discusión	32
CONCLUSIONES	33
RECOMENDACIONES	34
BIBLIOGRAFÍA	35
ANEXOS	42
Cronograma (Gantt)	42
Presupuesto Ejecutado	42

INDICE DE TABLAS

Tabla 1 Criterios de Inclusión y Exclusión.....	22
Tabla 2 Registro de la Búsqueda.....	23
Tabla 3 Nivel de Adopción del Aprendizaje Automático.....	27
Tabla 4 Herramientas del Aprendizaje Automático.....	29
Tabla 5 Limitaciones para aplicar el Aprendizaje Automático.....	31
Tabla 6 Presupuesto.....	42

INDICE DE FIGURAS

Figura 1 Línea de Investigación.....	9
Figura 2 Flujograma PRISMA.....	25
Figura 3 Cronograma de Gantt.....	42

INTRODUCCIÓN

El desarrollo de software se ha visto muy influenciado por los avances tecnológicos que tienen como objetivo optimizar y mejorar los procesos de desarrollo y mantenimiento de aplicaciones en la actual era digital. El aprendizaje automático (AA), un subcampo de la inteligencia artificial que permite a las máquinas aprender y mejorar a partir de datos sin instrucción explícita, es un área de interés reciente para investigadores y profesionales en este campo. Este método ha tenido mucho éxito a la hora de abordar problemas complejos y mejorar el trabajo manual en diferentes campos.

La historia del aprendizaje automático o también conocido como machine learning (ML) comienza en 1943 con el trabajo de Walter Pitts y Warren McCulloch, quienes modelaron matemáticamente las redes neuronales humanas. En 1950, Arthur Samuel desarrolló un programa de ordenador para jugar a las damas utilizando el algoritmo minimax y acuñó el término "machine learning" en 1952. En IBM, Frank Rosenblatt creó el Perceptrón para reconocimiento de imágenes, aunque inicialmente tuvo limitaciones (Fernández | Tokio, n.d., 2024).

En los años 70, la IA y el AA tomaron caminos separados, cuando la IA se centró en enfoques lógicos y el AA se orientó hacia la teoría de la probabilidad y estadísticas. El crecimiento de Internet y la disponibilidad de grandes cantidades de datos impulsaron el desarrollo de nuevos

algoritmos y aplicaciones de AA, desde el reconocimiento de voz hasta el ajedrez (Fernández | Tokio, n.d., 2024).

Hoy en día, el aprendizaje automático es crucial en diversos sectores, mejorando el rendimiento de sistemas informáticos y beneficiándose del desarrollo de Big Data. Las aplicaciones actuales incluyen vehículos autónomos, robots, automatización de procesos, análisis, y chatbots, entre otros. El aprendizaje automático continúa creciendo y ofrece oportunidades significativas para quienes se capaciten en esta disciplina (Fernández | Tokio, n.d., 2024).

El aprendizaje automático se refiere al desarrollo de sistemas que pueden aprender y mejorar su rendimiento a partir de los datos que procesan. Aunque los términos aprendizaje automático (AA) e inteligencia artificial (IA) suelen mencionarse juntos y a menudo se utilizan indistintamente, en realidad no significan lo mismo. Es importante destacar que, aunque el aprendizaje automático es un tipo de IA, no todas las aplicaciones de la inteligencia artificial dependen de este enfoque (Oracle, 2024).

El aprendizaje automático (AA) suele ser confundido muchas veces con la inteligencia artificial (IA), pero existe una diferencia fundamental que ayuda a identificar la función de cada una, la IA abarca diversas técnicas para hacer que las máquinas razonen, aprendan y resuelvan problemas como los humanos, mientras tanto, el AA es una técnica específica dentro de la IA que desarrolla algoritmos y modelos estadísticos para que las máquinas realicen tareas complejas sin instrucciones explícitas; en lugar de seguir reglas codificadas manualmente, las máquinas aprenden de los datos, identificando patrones e inferencias que les permiten mejorar su

rendimiento y adaptarse a nuevas situaciones, en resumen, la diferencia entre IA y AA radica en que la IA es un campo amplio que engloba diversas técnicas para que las máquinas resuelvan problemas de manera autónoma, mientras que el AA es una técnica específica dentro de la IA que se centra en mejorar esas capacidades mediante el aprendizaje a partir de datos (UNIE, n.d., 2023).

El presente trabajo de investigación tiene como objetivo realizar una revisión exhaustiva de la literatura existente sobre la aplicación del aprendizaje automático en el desarrollo de software. Se explorarán los diversos enfoques y técnicas que han sido implementados, así como los beneficios, desafíos y oportunidades que esta intersección ofrece. A través de esta revisión, se pretende proporcionar una comprensión clara y detallada del estado actual de la investigación en esta área, identificando tanto las tendencias emergentes como las brechas de conocimiento que aún persisten.

Al examinar profundamente la literatura disponible, esta investigación no solo busca proporcionar una visión integral de las aplicaciones prácticas del aprendizaje automático en el desarrollo de software, sino también ofrecer conocimientos para futuras investigaciones y prácticas en la industria.

RESUMEN

El desarrollo de software se enfrenta a una evolución constante debido al impacto de nuevas tecnologías, entre ellas el aprendizaje automático, que ha demostrado un gran potencial para mejorar la eficiencia y precisión en diversas etapas del proceso. A medida que la industria del software continúa avanzando hacia una mayor automatización, es crucial evaluar cómo se están adoptando estas técnicas y cuáles son las principales limitaciones que conllevan. Este trabajo de investigación tiene como objetivo determinar el estado actual de la aplicación del aprendizaje automático en el desarrollo de software, sintetizando los métodos utilizados y formulando recomendaciones para investigaciones futuras, se llevó a cabo una revisión sistemática de la literatura utilizando la metodología PRISMA, la cual permitió garantizar una visión precisa del conocimiento disponible. Los principales resultados indicaron un aumento significativo en la adopción del aprendizaje automático, especialmente en la predicción de fallos, optimización de código y detección de errores, siendo las técnicas más utilizadas las Máquinas de Vectores de Soporte (SVM), Redes Neuronales Convolucionales (CNN) y algoritmos de agrupamiento no supervisado como K-means. Sin embargo, se identificaron limitaciones relacionadas con la calidad de los datos y la necesidad de recursos computacionales elevados. Se concluye que, aunque el aprendizaje automático está transformando el desarrollo de software, los desafíos técnicos y éticos, como la dependencia de datos y los recursos, deben ser abordados para maximizar su impacto. Por tanto, se recomienda continuar investigando enfoques que optimicen su implementación, garantizando la equidad y transparencia en los sistemas desarrollados.

Palabras clave: aprendizaje automático, desarrollo de software, revisión sistemática, PRISMA, redes neuronales.

ABSTRACT

Software development faces constant evolution due to the impact of new technologies, particularly machine learning, which has shown significant potential to improve efficiency and accuracy in various stages of the process. As the software industry continues to move towards greater automation, it is crucial to evaluate how these techniques are being adopted and what the main limitations are. This research aims to determine the current state of machine learning application in software development, synthesizing the methods used and formulating recommendations for future research. A systematic literature review was conducted using the PRISMA methodology, which ensured an accurate view of the available knowledge. The main results indicated a significant increase in the adoption of machine learning, especially in failure prediction, code optimization, and error detection, with the most commonly used techniques being Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and unsupervised clustering algorithms like K-means. However, limitations related to data quality and the need for high computational resources were identified. It is concluded that although machine learning is transforming software development, technical and ethical challenges -such as dependence on data and resources- must be addressed to maximize its impact. Therefore, it is recommended to continue researching approaches that optimize its implementation, ensuring fairness and transparency in the developed systems.

Keywords: machine learning, software development, systematic review, PRISMA, neural networks.

CAPÍTULO I

FORMULACIÓN GENERAL DEL PROYECTO

Descripción del Problema

El desarrollo de software ha experimentado una transformación significativa en las últimas décadas, impulsada en gran medida por los avances en inteligencia artificial y aprendizaje automático. La capacidad de las máquinas para aprender de los datos y mejorar su rendimiento en tareas específicas ha llevado a la integración de técnicas de aprendizaje automático en numerosos aspectos del ciclo de vida del desarrollo de software. Sin embargo, a pesar del creciente interés y la adopción de estas tecnologías, existen desafíos y preguntas importantes que requieren una investigación más profunda.

Las causas de esta transformación se encuentran en la necesidad de mejorar la eficiencia y efectividad en el desarrollo de software, la creciente complejidad de los sistemas tanto en aplicaciones, y la disponibilidad de grandes cantidades de datos que pueden usarse para entrenar modelos de aprendizaje automático. Además, la competencia en el mercado tecnológico obliga a las empresas a implementar soluciones innovadoras que permitan la optimización de procesos y productos.

Las consecuencias de la integración del aprendizaje automático en el desarrollo de software son variadas y significativas. Por un lado, hay una mejora en la precisión y velocidad de tareas como la detección de errores, la predicción de fallos y la optimización del código. Esto puede ayudar a reducir los costos y el tiempo de desarrollo, así como a obtener un software de mayor calidad y más seguro. Por otro lado, también existen desafíos como la necesidad de contar con personal especializado en técnicas de aprendizaje automático, la dependencia de datos de alta calidad y la gestión de los riesgos que implica el despliegue de modelos que pueden ser complejos y difíciles de explicar.

(Para et al., 2020) en su trabajo de tesis “Recocido simulado y aprendizaje automático aplicado a la optimización de factores de riesgo empresarial” utiliza la técnica de K-means para particionar datos en clústeres sin etiquetas predefinidas, identifica que su principal limitación es la convergencia no garantizada y la dependencia de la calidad de los datos.

(Valdivieso et al., 2019) en su trabajo de tesis “Mhealth con big data y machine learning como soporte tecnológico para la salud en Colombia” se enfoca en sistemas de recomendación que analizan información histórica de usuarios para ofrecer opciones personalizadas en el contexto de la salud, identifica que su principal limitación es la dependencia de la calidad de los datos y el manejo de datos atípicos.

En esta investigación se centrará en comprender el estado actual del aprendizaje automático aplicado al desarrollo de software, se busca explorar y describir qué técnicas de aprendizaje automático se están utilizando, cómo se están implementando en la práctica, qué tipos de problemas se están abordando y cuáles son los resultados obtenidos.

Formulación del Problema

¿Cuál es el estado actual de la aplicación del aprendizaje automático en el proceso de desarrollo de software?

Preguntas de Investigación

¿Cuál es el nivel de adopción del aprendizaje automático para el desarrollo de software?

¿De qué manera se ha aplicado el aprendizaje automático en el proceso de desarrollo de software?

¿Cuáles son las limitaciones para la adopción del aprendizaje automático en el proceso de desarrollo de software?

Justificación

Esta investigación es fundamental debido a la creciente importancia del desarrollo de software en la sociedad actual y al papel crucial que desempeña el aprendizaje automático en este campo. La situación problemática que se pretende investigar radica en que, en un entorno donde el desarrollo de software es un pilar esencial para múltiples industrias, desde la salud hasta las finanzas, es fundamental comprender cómo el aprendizaje automático puede mejorar la eficiencia, confiabilidad y adaptabilidad del software. La complejidad de los sistemas de software modernos exige soluciones avanzadas que solo el aprendizaje automático puede ofrecer. La relevancia tecnológica del aprendizaje automático es una de las principales motivaciones para esta investigación. Con la rápida evolución de las tecnologías, la integración del aprendizaje automático en el desarrollo de software no solo es pertinente, sino necesaria para mantener la competitividad en el mercado actual. Este estudio se centrará en explorar y describir las técnicas de aprendizaje automático utilizadas en el desarrollo de software, analizando cómo se implementan en la práctica, los tipos de problemas que abordan y los resultados obtenidos, así como las causas de su adopción y las consecuencias de su implementación en la industria.

Los beneficios sociales de este trabajo radican en que la comprensión crítica de cómo el aprendizaje automático transforma el desarrollo de software permitirá a organizaciones y profesionales del sector tomar decisiones informadas y estratégicas sobre la integración de estas tecnologías. Esto no solo beneficiará a las empresas, sino también a los usuarios finales, al ofrecer productos de software más eficientes y de mayor calidad.

Finalmente, este trabajo contribuirá al conocimiento existente sobre el aprendizaje automático y su aplicación en el desarrollo de software, proporcionando un marco de referencia para futuras investigaciones. Además, establecerá metodologías que podrían ser utilizadas por otros investigadores en el campo, fomentando un enfoque más disciplinado en el estudio de estas tecnologías emergentes. En resumen, esta investigación debe llevarse a cabo para ofrecer una comprensión profunda de cómo el aprendizaje automático está transformando el desarrollo de software, identificando sus beneficios y desafíos, lo que permitirá a los profesionales del sector adaptarse a las demandas de la era actual, donde el aprendizaje automático tiene la capacidad de potenciar significativamente el desarrollo de software.

Línea de Investigación.

Figura 1 Línea de Investigación

Dominio	Línea	
Tecnologías de la Información y Comunicación	Gestión De Tecnologías de la Información y Comunicación	
	Ingeniería De Software, Redes y Telecomunicaciones	x
	Educación Virtual, Teleeducación	
	Big Data, Cloud Computing, Gestión de Datos e Inteligencia Artificial	
	Geomática	

Sublínea	Modelos de desarrollo de software
----------	-----------------------------------

Objetivos: General y Específicos

General

Determinar el estado actual de la aplicación del aprendizaje automático en el desarrollo de software, sintetizando los métodos utilizados y formulando recomendaciones para futuras investigaciones y aplicaciones.

Específicos

- Evaluar el nivel de adopción del aprendizaje automático para el desarrollo de software.
- Identificar los principales métodos, técnicas y aplicaciones del aprendizaje automático en el proceso de desarrollo de software.
- Sintetizar las principales limitaciones para la adopción del aprendizaje automático en el proceso de desarrollo de software.

Idea a Defender

Existe una relación entre el aprendizaje automático y el desarrollo de software.

Variables

No existe variables, porque se va a realizar una revisión de literatura, no se va a establecer una correlación entre variables y tampoco una relación de causalidad, es una investigación donde se realizará un estudio descriptivo.

CAPÍTULO II MARCO TEÓRICO

Antecedentes

De acuerdo con Salas en su tesis de grado Algoritmos de aprendizaje automático para predecir la curación de malformaciones arteriovenosas cerebrales tratadas con radiocirugía, tuvo como objetivo identificar factores asociados a la curación de las MAVs cerebrales en pacientes tratados con radiocirugía estereotáctica y comparar algoritmos de aprendizaje automático para predecirla. Se analizaron los datos de 202 pacientes tratados en el período 2005-2018 con 45 variables clínicas, sociodemográficas, morfológicas y hemodinámicas. El resultado fue un total de 83 pacientes logró la curación en el seguimiento a largo plazo, con una frecuencia de los siguientes síntomas: cefalea – 88.12%, hemorragia cerebral – 76.73% y convulsiones – 55%. Los algoritmos de máquina de soporte vectorial, Random Forest y Redes Neuronales se evaluaron en un conjunto de datos que supone el 80% de los datos para entrenamiento y el 20% de los datos para pruebas. La SVM mostró el mejor rendimiento con una precisión del 98.51%, una sensibilidad del 100% y una especificidad del 96.97%. Si los algoritmos se introducen en las mismas guías para prever la curación de las MAV, se pueden adoptar decisiones mucho más personalizadas y efectivas (Salas, 2021).

Según Parrales y Rosales en su trabajo de titulación Utilización de Machine Learning para el proceso de selección de personal en una microempresa, tiene como objetivo analizar cómo se ejecuta el reclutamiento de personal mediante un algoritmo integrado en un sistema informático. En términos más simples, este método recopila datos y luego utiliza un patrón para elegir a las personas más adecuadas para un trabajo.

La investigación utilizó un método simple, no numérico, principalmente mediante cuestionarios, para recopilar información. Se encuestaron a 52 personas, con edades comprendidas entre 18 y 46 años, en su mayoría provenientes de Guayaquil, Quito y Cuenca. Los participantes trabajan en empresas públicas y privadas, elegidos debido a su uso predominante de tácticas de selección basados en Machine Learning o la creación de estos algoritmos para entidades externas.

Los resultados establecieron que la implementación de inteligencia artificial y aprendizaje automático mejora sustancialmente los métodos de evaluación de candidatos, brindando ventajas favorables a las corporaciones (Parrales & Rosales, 2022).

En el artículo Machine Learning Aplicado al Análisis del Rendimiento de Desarrollo de Software (Gil & Seguro, 2022), buscaban identificar la metodología más efectiva para predecir si un software cumple con las necesidades del cliente. Utilizando datos de pruebas de rendimiento y la métrica F1-score, se encontró que Random Forest tuvo el mejor desempeño. Sin embargo, la investigación subrayó que la calidad y cantidad de datos, así como un buen procesamiento, son fundamentales para obtener resultados confiables, independientemente del método empleado.

Según Paico y Valdera en su tesis de grado Detección automática de ataques de inyección mediante aprendizaje automático en bases de datos NoSQL, aborda la creciente popularidad de las bases de datos NoSQL, impulsada por tecnologías como Big Data e IoT.

Si bien estas son bases para operar con información en tiempo real a la vez, los riesgos de seguridad a los que están sometidas estas bases son críticos y aún subestimados, especialmente los Ataques de Inyección. El estudio desarrolló un modelo basado en aprendizaje automático sobre bases de datos NoSQL basadas en documentos para detectar ataques a esta información. Se utilizó un set de datos de 19 atributos y 509 ejemplos generados a través de estructuras JSON, que, a través de algoritmos como Redes Neuronales, Árboles de Decisión y el Random Forest obtuvieron una exactitud muy precisa. Donde las Redes Neuronales y el Perceptrón Multicapa estuvieron en un 98.7 %. El escenario de trabajo promueve que estos modelos pueden ser utilizados para detectar ataques de forma eficiente. El futuro sería crear un Dataset que involucre otro tipo de bases NoSQL (Paico & Valdera, 2021).

Los trabajos ya antes mencionados (artículos, tesis) permiten fortalecer mi proyecto de investigación, al aportarme información sobre el tema ya antes mencionado, se podrá analizar cuál sería el estado actual de la aplicación del aprendizaje automático en el proceso de desarrollo de software, y poder determinar sus principales limitaciones y aplicaciones que el aprendizaje automático tiene con el desarrollo del software.

Marco Científico

Aprendizaje Automático (AA).

El aprendizaje automático es un subconjunto específico de IA, refiere a darle a la máquina un ciclo de retroalimentación que le permite aprender de la experiencia. Como término, el aprendizaje automático solo existe desde la década de 1980. Recientemente, en los últimos 10-15 años, se ha tenido el poder de procesamiento y almacenamiento de datos para realmente comenzar a implementar el AA a escala (T. Norman, 2019).

Es un tipo de inteligencia artificial que permite a las máquinas aprender de los datos sin necesidad de programarlas explícitamente. Para ello, optimiza los parámetros del modelo (es decir, variables internas) mediante cálculos, de manera que el comportamiento del modelo refleje los datos o la experiencia. Como paso siguiente, el algoritmo de aprendizaje actualiza continuamente los valores de los parámetros a medida que el aprendizaje progresa, lo que permite al modelo de AA aprender y hacer predicciones o tomar decisiones basadas en la ciencia de datos (ISO, 2024).

Evolución del Aprendizaje Automático.

La historia del machine learning comienza a mediados del siglo XX y se trata de una disciplina que se concibe por primera vez a partir del modelado matemático de las redes neuronales humanas. Walter Pitts y Warren McCulloch publican en 1943 un artículo que intenta mapear matemáticamente el pensamiento y la toma de decisiones en los seres humanos.

Unos años después de este primer hito, en 1950, Arthur Samuel desarrolla un programa de ordenador para jugar a las damas. En este programa, con una memoria muy limitada, la máquina es capaz de elegir sus movimientos mediante la estrategia minimax (tomar decisiones esperando el peor escenario en la respuesta del adversario), que pasó a denominarse algoritmo minimax.

Samuel diseñó mecanismos que mejoraban sus primeras versiones del programa y lo llamó aprendizaje de memoria. El programa que había desarrollado fue capaz de registrar y memorizar todas las posiciones que había visto y al combinar esto con lo que ya había, el juego de las damas

mejoró. Ya en 1952, al científico se le ocurrió el nombre de machine learning (Fernández | Tokio, n.d., 2024).

En sus inicios, el desarrollo de la investigación en machine learning y IA avanzaba juntos. Sin embargo, hacia finales de los años 70, la IA se orientó hacia enfoques lógicos basados en el conocimiento en lugar de algoritmos, distanciándose del machine learning. Además, la investigación en redes neuronales fue dejada de lado. Esto provocó una ruptura entre ambas disciplinas, y pasaron varios años antes de que el machine learning y las redes neuronales volvieran a tener impacto, resurgiendo en la década de 1990 (Fernández | Tokio, n.d., 2024).

La historia del machine learning experimentó un cambio significativo hacia la practicidad y la resolución de problemas reales en la industria. Este cambio se reflejó en la transición de enfoques heredados de la IA hacia el uso de la teoría de la probabilidad y las estadísticas. El crecimiento de Internet y la disponibilidad de grandes cantidades de datos digitales impulsaron aún más este cambio. Surgieron nuevos algoritmos y programas inteligentes capaces de una variedad de tareas, desde la planificación de rutas hasta el reconocimiento de voz y el dominio en juegos como el ajedrez (Fernández | Tokio, n.d., 2024).

En la actualidad, el machine learning se ha vuelto imprescindible en diversos sectores empresariales y de investigación. La combinación de algoritmos de aprendizaje automático con modelos de redes neuronales está potenciando el rendimiento de los sistemas informáticos. Además, el auge de Internet ha generado enormes volúmenes de datos, impulsando el desarrollo del Big Data y su integración en procesos empresariales, lo que ha mejorado aún más los algoritmos y el avance del aprendizaje automático (Fernández | Tokio, n.d., 2024).

Técnicas de Aprendizaje Automático.

Aprendizaje supervisado:

El aprendizaje supervisado es un subconjunto del machine learning que consiste en la deducción de información a partir de datos de entrenamiento. Estos datos se clasifican en dos secciones: datos de entrenamiento y datos de prueba.

Los datos de entrenamiento se utilizan para entrenar a un modelo, y los datos de prueba son los que se usan para determinar la eficacia del modelo creado.

El objetivo del aprendizaje supervisado es crear un programa que sea capaz de resolver cualquier variable de entrada luego de ser sometido a un proceso de entrenamiento (De Ceupe, n.d., 2022).

Aprendizaje no supervisado:

El aprendizaje no supervisado busca dejar que la máquina tome las decisiones y actúe por libre. El quid de la cuestión radica en que no se le dice al ordenador qué tipo de datos son los que debe predecir.

En este tipo de aprendizaje los algoritmos no producen etiquetas para los datos, ni siquiera interesan los datos de entrada o de salida.

Aquí entra el concepto de clustering, muy usado en la minería de datos para referirse a una técnica para identificar automáticamente grupos objetivo y aglutinarlos en un mismo conjunto según ciertas similitudes (Navarro, n.d., 2024).

Aprendizaje por refuerzo:

El aprendizaje por refuerzo es una rama del machine learning en la cual la máquina guía su propio aprendizaje a través de recompensas y castigos. Es decir, consiste en un sistema de instrucción autónomo cuyo camino es indicado según sus aciertos y errores.

Consta de un aprendizaje empírico, por lo que el agente informático está en constante búsqueda de aquellas decisiones que le premien de algún modo, a la par que evita aquellos caminos que, por experiencia propia, son penalizados.

El aprendizaje reforzado es un concepto similar al que utilizan los seres vivos. Esto es, las máquinas aprenden qué decisiones tomar de acuerdo a la situación en la que se encuentren. Además, son capaces de desarrollar estrategias con una visión a largo plazo (Natalia, n.d., 2022).

Aprendizaje semi supervisado:

El aprendizaje semisupervisado es una técnica de machine learning que combina el aprendizaje supervisado y no supervisado para aprovechar conjuntos de datos que contienen pocos ejemplos etiquetados y muchos ejemplos no etiquetados, se utilizan algoritmos de aprendizaje no supervisado para extraer características relevantes y representaciones útiles de los datos no etiquetados, y luego se utilizan estos conocimientos para mejorar la calidad del modelo de aprendizaje supervisado.

Este aprendizaje es particularmente útil en aplicaciones donde la recopilación de datos etiquetados es costosa o difícil, pero donde se dispone de una gran cantidad de datos no etiquetados. Se ha demostrado que el aprendizaje semisupervisado mejora significativamente la precisión de los modelos de machine learning en aplicaciones de reconocimiento de voz, visión por computadora y procesamiento de lenguaje natural (Gamco, 2024).

Aprendizaje profundo (Deep learning):

El aprendizaje profundo es una rama del aprendizaje automático y la inteligencia artificial que imita la adquisición de conocimiento de los humanos. Es crucial en la ciencia de datos, facilitando la recolección, análisis e interpretación de grandes volúmenes de datos. A diferencia de los algoritmos tradicionales de aprendizaje automático que son lineales, los algoritmos de aprendizaje profundo tienen una estructura jerárquica de mayor complejidad y abstracción, lo que automatiza el análisis predictivo (Burns, n.d., 2021).

Desarrollo de software.

Es el proceso de crear y diseñar programas o aplicaciones informáticas mediante la escritura de código, diseño, pruebas y mantenimiento. Este proceso implica una serie de pasos e instrucciones que los desarrolladores siguen para crear software funcional y de alta calidad. Se trata de todos los procesos involucrados para el desarrollo, desde el origen, la producción, revisión, la entrega y posterior mantenimiento (Adminstark, n.d., 2024).

Evolución del desarrollo de software.

En las últimas dos décadas, el desarrollo de software ha cambiado drásticamente. Anteriormente, se usaban lenguajes como Pascal en MS-DOS, con código generado mayormente por los programadores y almacenado en disquetes. Los proyectos se seguían de forma secuencial, con mucho trabajo manual en pruebas, distribución y soporte. Hoy en día, se emplean lenguajes híbridos que combinan paradigmas como la orientación a objetos, la programación funcional y la reactiva. Gran parte del código ahora proviene de librerías open source, Stack Overflow y asistentes de código inteligentes.

El desarrollo de software es continuo, con nuevas versiones lanzadas semanalmente o incluso a diario, gracias a la automatización de muchos procesos. Recientemente, la IA generativa ha comenzado a automatizar el diseño y la codificación de aplicaciones, acelerando aún más el desarrollo. Las empresas ya no necesitan sus propios servidores, ya que la gestión de infraestructura se externaliza a la nube, permitiendo menores barreras de entrada y economías de escala.

Sin embargo, aunque metodologías como Agile, Devops, Microservicios, el Cloud y el SaaS han traído grandes beneficios, también presentan desafíos que a menudo se ignoran. Es esencial entender la filosofía detrás de estas nuevas herramientas y mantener una mirada crítica para ser conscientes de los nuevos retos. Al adoptar la IA en el desarrollo de software, es fundamental usar las experiencias pasadas para no dejar que sus beneficios oculten los inconvenientes (Di Pierro, n.d., 2024).

Metodología Prisma.

PRISMA es el acrónimo de Preferred Reporting Items for Systematic reviews and Meta-Analyses y fue desarrollada por un grupo de expertos en metodología y epidemiología, para ofrecer una guía a la hora de asegurar la transparencia de las revisiones sistemáticas. El método PRISMA o declaración PRISMA es una guía publicada en el año 2009 para ayudar a los investigadores que necesiten hacer una revisión sistemática a documentar, con transparencia, qué hicieron otros autores y el porqué de su propia revisión. Esta metodología te proporciona cierta seguridad cuando presentas una revisión sistemática sola o como apartado de una tesis, por su aceptación y por el hecho de ser cada vez más utilizada. Se puede aplicar, también, a publicaciones en revistas especializadas. El uso del método PRISMA ayuda a garantizar una revisión sistemática de alta calidad. Puede parecer trabajoso, pero ha sido desarrollado para asegurar el rigor en los resultados y agilizar el trabajo que supone hacer una revisión sistemática, su amplia aceptación es una garantía, sobre todo si te preocupa obviar sesgos o presentar la información de un modo poco estructurado. Creemos que este puede ser el método más habitual en un corto espacio de tiempo, por lo que te conviene estar familiarizado con él. (Tesis Doctorales Online, 2024).

Marco Conceptual

Aprendizaje Automático o Machine Learning (AA/ML):

Es una rama específica dedicada a ayudar a las computadoras a aprender de los humanos y cómo interactuar con nosotros de una manera similar a la de los humanos (Torres, n.d., 2019).

Desarrollo de Software:

El desarrollo de software es el proceso de crear y diseñar programas o aplicaciones informáticas mediante la escritura de código, diseño, pruebas y mantenimiento (Adminstark, n.d., 2024).

Marco Legal

“Código Orgánico de la economía social de los conocimientos, creatividad e innovación”

Título Preliminar

De las disposiciones comunes a la economía social de los conocimientos, creatividad e innovación.

Art 4.- Principios. –

Los derechos intelectuales son una herramienta para la adecuada gestión de los conocimientos. La adquisición y ejercicio de los derechos de propiedad intelectual asegurarán un equilibrio entre titulares y usuarios. Además de las limitaciones y excepciones previstas en este Código, el Estado podrá adoptar las medidas necesarias para garantizar la salud, nutrición, educación, cultura, el desarrollo científico y tecnológico, la innovación y la transferencia y difusión tecnológica como sectores de importancia vital para el desarrollo socioeconómico y tecnológico del país. Nada de lo previsto en este Código podrá interpretarse de forma contraria a los principios, derechos y obligaciones establecidos en los Tratados Internacionales de los que Ecuador es parte, como parte integrante de nuestro ordenamiento jurídico.

Libro III

De la Gestión de los Conocimientos

Título I

Principios y Disposiciones Generales

Capítulo I

Principios Generales

Art. 85.- Derechos intelectuales. - Se protegen los derechos intelectuales en todas sus formas, los mismos que serán adquiridos de conformidad con la Constitución, los Tratados Internacionales de los cuales Ecuador es parte y el presente Código. Los derechos intelectuales comprenden principalmente a la propiedad intelectual, y los conocimientos tradicionales. Su regulación constituye una herramienta para la adecuada gestión de los conocimientos, con el objetivo de promover el desarrollo científico, tecnológico, artístico, y cultural, así como para incentivar la innovación. Su adquisición y ejercicio, así como su ponderación con otros derechos, asegurarán el efectivo goce de los derechos fundamentales y contribuirán a una adecuada difusión de los conocimientos en beneficio de los titulares y la sociedad.

CAPITULO III METODOLOGÍA

Tipo de Investigación

Investigación descriptiva

La investigación descriptiva es una herramienta esencial en la ciencia que se enfoca en observar y describir las características de grupos de personas, objetos o eventos. Su objetivo principal es ofrecer un retrato detallado de la situación tal como es en el momento del estudio, sin investigar las causas. Esta investigación recopila y organiza datos para proporcionar una imagen clara de lo que ocurre, empleando técnicas como la clasificación y el resumen, y utilizando medidas estadísticas como la media y la variabilidad para comprender mejor el tema estudiado (Arias, n.d., 2024).

Este tipo de investigación nos permitirá obtener una visión detallada y estructurada sobre cómo se está utilizando el aprendizaje automático (AA) en el campo del desarrollo de software, además, permitirá obtener información valiosa tanto para investigadores como para profesionales en el campo.

Enfoque de la investigación

La investigación que se realizará va a tener un enfoque cualitativo, lo que permitirá una comprensión profunda y detallada del uso del aprendizaje automático en el desarrollo de software. Este tipo de estudio facilitará captar la complejidad del tema y ofrecer perspectivas valiosas que serían difíciles de lograr con un análisis exclusivamente cuantitativo.

Métodos de Investigación

Método inductivo

Se empleará el método inductivo, ya que permitirá explorar en detalle el uso del aprendizaje automático en el desarrollo de software. Este enfoque facilitará la recolección de datos específicos y la identificación de patrones y tendencias.

Técnicas e Instrumentos de Recopilación de Datos

Bibliográfica/Documental

La revisión bibliográfica permitirá acceder a una amplia variedad de fuentes académicas y técnicas, como revistas especializadas y libros, facilitando una exploración exhaustiva de la literatura existente.

El análisis documental, por su parte nos permitirá analizar documentos técnicos relevantes, como informes de investigación (artículos) y proyectos prácticos donde se ha aplicado el aprendizaje automático en el desarrollo de software.

Procesamiento de la Información

Para el procesamiento de la información se va a utilizar la metodología PRISMA, ya que cuenta con un marco estructurado y claro para realizar revisiones sistemáticas de la literatura, esta metodología nos garantiza una visión completa y precisa del conocimiento actual sobre el tema, minimizando sesgos, y brindando una mayor efectividad del proceso de revisión.

CAPITULO IV RESULTADOS Y DISCUSIÓN

Análisis, Interpretación y Discusión de Resultados

Metodología PRISMA.

1. Desarrollo del Protocolo de Revisión:

Definir la Pregunta de Investigación:

¿Cuál es el estado actual de la adopción del aprendizaje automático en el desarrollo de software, considerando sus aplicaciones, métodos empleados y limitaciones?

Criterios de inclusión y de exclusión

Tabla 1 Criterios de Inclusión y Exclusión

Criterios de inclusión	Criterios de exclusión
<ul style="list-style-type: none"> ▪ Artículos que traten el tema planteado para esta investigación. ▪ Artículos de hasta 5 años de antigüedad. ▪ Artículos de bases de datos o repositorios de acceso gratuito. ▪ Artículos en español o en inglés que traten el tema planteado y sean aplicables en el contexto latinoamericano. ▪ Artículos generados en el contexto latinoamericano. 	<ul style="list-style-type: none"> ▪ Artículos que no traten el tema planteado para esta investigación. ▪ Artículos de más de 5 años de antigüedad. ▪ Artículos de bases de datos o repositorios que sean pagados. ▪ Artículos en español o en inglés que no traten el tema planteado y no sean aplicables en el contexto latinoamericano. ▪ Artículos generados fuera del contexto latinoamericano.

2. Estrategia de Búsqueda:

La estrategia de búsqueda a seguir para la investigación en base de datos se basa en utilizar palabras claves pertenecientes al tema. Ejemplo:

Scielo (“Aprendizaje automático” AND “Desarrollo de software”) AND year_cluster (“2020” AND “2024”).

Registro de la Búsqueda:

Registrar detalles de la búsqueda, incluidas las bases de datos utilizadas y términos de búsqueda.

Tabla 2 Registro de la Búsqueda

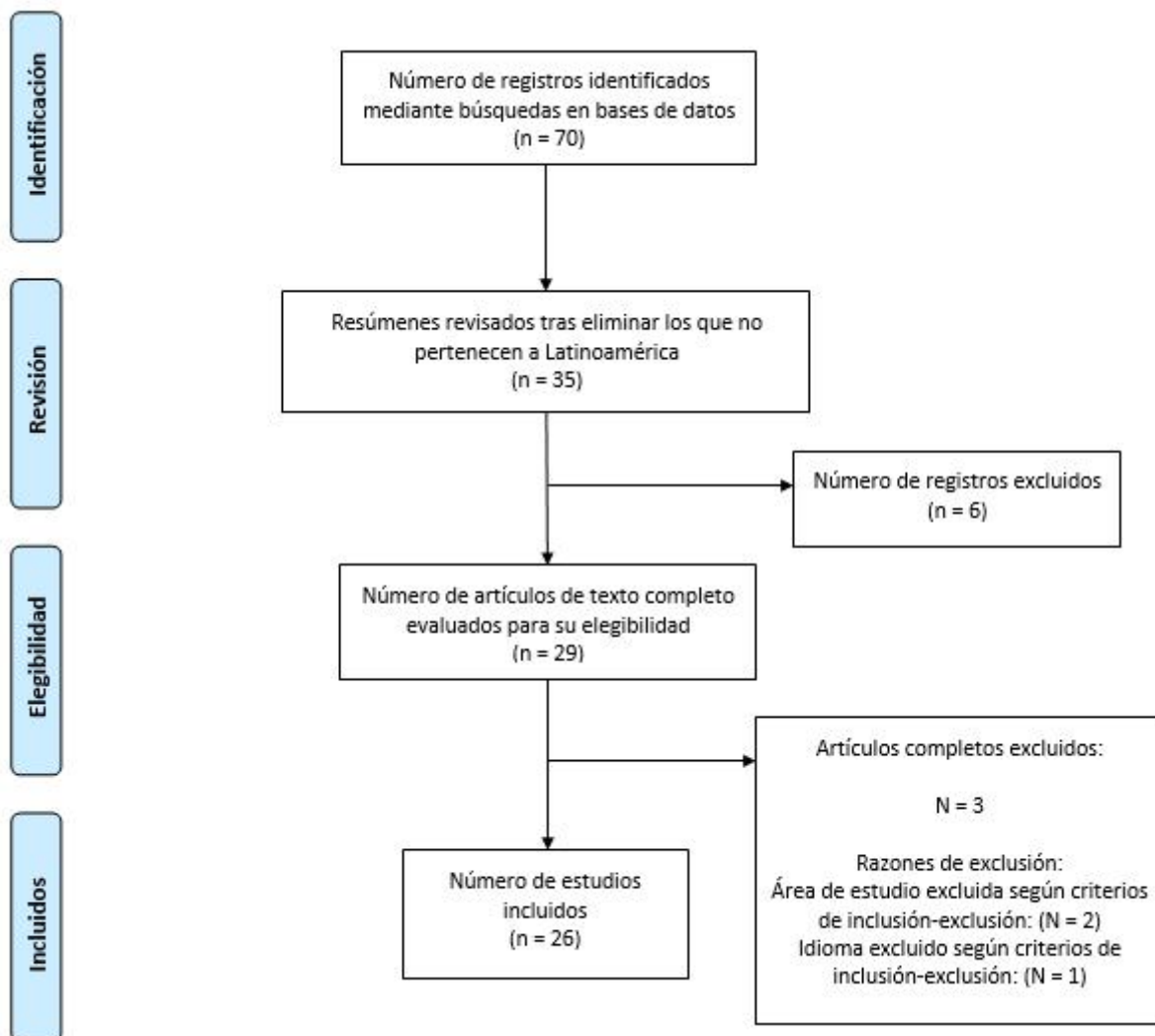
Base de datos	Palabras de búsqueda
Scopus	(Machine learning) AND (Software development) AND PUBYEAR (2020 AND 2024)
Web of Science	“Aprendizaje automático” AND “Desarrollo de software” AND PUBYEAR “2020 AND 2024”
Scielo	(“Aprendizaje automático” AND “Desarrollo de software”) AND year_cluster (“2020” AND “2024”).
Redalyc	(“Aprendizaje automático” AND “Desarrollo de software”) AND year_cluster (“2020” AND “2024”).
Google Académico	(“Aprendizaje automático” AND “Desarrollo de software”) AND year_cluster (“2020” AND “2024”).
DOAJ	(“Aprendizaje automático” AND “Desarrollo de software”) AND year_cluster (“2020” AND “2024”).
Dialnet	(“Aprendizaje automático” AND “Desarrollo de software”) AND year_cluster (“2020” AND “2024”).

3. Selección de Estudios:

Se realizó una búsqueda de literatura en siete bases de datos: Scopus, Web of Science, Dialnet, DOAJ, Google Académico, Redalyc y Scielo. La búsqueda inicial identificó un total de 70 artículos. Después de eliminar duplicados y aplicar los criterios de inclusión y exclusión, 29 artículos fueron evaluados en texto completo. Finalmente, 26 artículos cumplieron con todos los criterios de inclusión y fueron incluidos en la síntesis cualitativa.

El proceso de búsqueda y selección se ilustra en el flujograma PRISMA, que muestra el número de artículos identificados, seleccionados, excluidos y finalmente incluidos en la revisión.

Figura 2 Flujograma PRISMA



Análisis y Síntesis de los Resultados.

¿Cuál es el nivel de adopción del aprendizaje automático para el desarrollo de software?

La adopción del aprendizaje automático (AA) en el desarrollo de software ha experimentado un crecimiento significativo, extendiéndose a diversos dominios, desde la salud y la agricultura hasta el comercio electrónico y el análisis de sentimientos. La revisión de los artículos muestra que el AA se está utilizando de manera diversa y extensa, lo que destaca su potencial para transformar múltiples áreas del desarrollo de software.

En el ámbito del agrupamiento y clasificación de datos, el uso de técnicas de aprendizaje no supervisado, como k-means y fuzzy c-means, es común para identificar patrones y clasificar información sin etiquetas previas. Estos métodos se aplican en el análisis de grandes volúmenes de datos, donde la identificación de patrones y la clasificación sin etiquetas son cruciales. La adopción de estas técnicas es alta, especialmente en investigación y análisis de datos, como se observa en la "Metodología de desarrollo de técnicas de agrupamiento de datos usando aprendizaje automático" (Contreras Contreras et al., 2022).

En el sector salud, el AA ha mostrado una adopción significativa. Sistemas de recomendación que analizan información/historial de usuarios para ofrecer opciones personalizadas, como se describe en "Mhealth con big data y machine learning como soporte tecnológico para la salud en Colombia," están optimizando la atención médica y prediciendo patrones de salud. (Valdivieso et al., 2019) . Además, técnicas de AA se están utilizando para la detección temprana de enfermedades, como la predicción de diabetes y arritmias, mejorando así los resultados de salud de los pacientes (Mejía et al., 2023).

La agricultura de precisión también se beneficia del AA, como se detalla en el "Diseño de un sistema para el control de riego mediante técnicas de aprendizaje automático aplicada a la agricultura de precisión en la granja La Pradera de la Universidad Técnica del Norte." Aquí, las redes neuronales se implementan para el control de riego en cultivos utilizando datos de sensores ambientales, lo que mejora la eficiencia y sostenibilidad de las prácticas agrícolas. La adopción en este campo es moderada pero creciente, a medida que los agricultores buscan formas más eficientes de gestionar sus cultivos (Cadena Lema, 2020).

El análisis de sentimientos es otra área donde el AA se está adoptando ampliamente. Técnicas de aprendizaje supervisado como SVM, Naive Bayes, y Random Forest se utilizan para clasificar sentimientos en un corpus de datos etiquetado con polaridades positivas y negativas, como se discute en "Aprendizaje automático aplicado al análisis de sentimientos." Esta adopción es alta en marketing, análisis de redes sociales y estudios de opinión pública, donde comprender las percepciones y emociones de los usuarios es fundamental (Cedeno-Moreno & Vargas-Lombardo, 2020).

En el ámbito educativo, el AA se está utilizando para predecir la deserción estudiantil, evaluando varios algoritmos como la regresión logística y los árboles de decisión. El artículo "El aprendizaje automático en entornos educativos universitarios: caso deserción académica" destaca cómo estas técnicas pueden ayudar a mejorar la retención de estudiantes en instituciones educativas. La adopción en este campo está en crecimiento, con instituciones cada vez más interesadas en utilizar el AA para mejorar sus tasas de retención estudiantil (Henríquez Miranda et al., 2022).

En el comercio electrónico, el AA se emplea para análisis predictivo y detección de spam, así como para personalizar las experiencias de los usuarios y optimizar las operaciones comerciales. La "Revisión de la literatura del comercio electrónico, el aprendizaje automático y sus aplicaciones en la industria y tiendas por departamento en línea" muestra cómo el AA está transformando la industria del comercio electrónico, mejorando la eficiencia operativa y la satisfacción del cliente. La adopción en este sector es alta, impulsada por la necesidad de mantenerse competitivos en un mercado digital en rápida evolución (Agama Espinoza, 2021).

Tabla 3 Nivel de Adopción del Aprendizaje Automático

N.º	Adopción	Referencia
1	Adopción en análisis de datos.	(Contreras Contreras et al., 2022)
2	Adopción en sistemas de recomendación.	(Valdivieso et al., 2019)
3	Adopción en control de sistemas autónomos.	(Cadena Lema, 2020)
4	Adopción en análisis de sentimientos.	(Cedeno-Moreno & Vargas-Lombardo, 2020)
5	Adopción en predicción de eventos.	(Henríquez Miranda et al., 2022)

6 Adopción en personalización de (Agama Espinoza, 2021) experiencias.

En conclusión, el aprendizaje automático es cada vez más frecuente en el desarrollo de software, abarcando campos como la salud, la agricultura, la educación y el comercio electrónico, su capacidad para analizar grandes volúmenes de datos, predecir patrones y optimizar procesos está transformando la forma en que se desarrollan y utilizan las aplicaciones de software, destacando su impacto y potencial en diversas áreas de estudio.

¿De qué manera se ha aplicado el aprendizaje automático en el proceso de desarrollo de software?

El aprendizaje automático se ha aplicado en el desarrollo de software para optimizar procesos, analizar datos masivos, automatizar tareas y personalizar soluciones en sectores como la salud, la agricultura, la educación y el comercio electrónico se usaron técnicas que incluyen algoritmos de aprendizaje supervisado, no supervisado y profundo.

En el aprendizaje supervisado, las Máquinas de Vectores de Soporte (SVM) y las Redes Neuronales Artificiales (ANN) son ampliamente usadas para clasificación y regresión, aplicadas en la predicción de enfermedades y sistemas de control automatizados, como el riego en agricultura (Cadena Lema, 2020). Las Redes Neuronales Convolucionales (CNN) se destacan en el procesamiento de imágenes, usadas para detectar mascarillas y clasificar objetos reciclables. (Chuquimarca Jimenez et al., 2021). Los Árboles de Decisión y Bosques Aleatorios son efectivos para predicción y clasificación en áreas como la predicción de diabetes y la recomendación de programas de capacitación (Mejía et al., 2023).

En el aprendizaje no supervisado, los algoritmos K-means y Fuzzy C-means se emplean para el agrupamiento de datos sin etiquetas, identificando patrones en grandes volúmenes de datos. (Contreras Contreras et al., 2022). El Análisis de Componentes Principales (PCA) es utilizado para la reducción de dimensionalidad y la identificación de clústeres en datos complejos (Flórez-Martíne et al., 2020).

El aprendizaje profundo, especialmente Deep Learning, utiliza redes neuronales profundas para tareas que requieren el procesamiento de grandes cantidades de datos y la extracción de características complejas, con aplicaciones en la detección de mascarillas para COVID-19 y la clasificación de manzanas mediante cámaras multiespectrales (Chuquimarca Jimenez et al., 2021).

Estas herramientas y técnicas permiten abordar una amplia gama de aplicaciones en el desarrollo de software, desde la salud y la agricultura hasta la gestión de datos y el comercio electrónico.

Tabla 4 Herramientas del Aprendizaje Automático

N.º	Herramientas	Referencia
1	Aprendizaje no supervisado (k-means y fuzzy c-means).	(Contreras Contreras et al., 2022)
2	SVM, random forest, redes neuronales artificiales.	(Valdivieso et al., 2019) (Polo-Triana et al., 2022) (Patiño et al., 2023)
3	Redes neuronales.	(Cadena Lema, 2020) (Flórez-Martínez et al., 2020) (Chuquimarca Jimenez et al., 2021)
4	SVM, Naive Bayes, Random Forest.	(Cedeno-Moreno & Vargas-Lombardo, 2020)
5	Regresión logística, árboles de decisión.	(Henríquez Miranda et al., 2022)
6	Árboles de decisión, redes neuronales.	(Agama Espinoza, 2021) (Mejía et al., 2023)

En conclusión, el aprendizaje automático ha demostrado ser una herramienta versátil y poderosa en el desarrollo de software, aplicándose en diversas industrias como la salud, la agricultura, la educación y el comercio electrónico. Las técnicas de aprendizaje supervisado, no supervisado y

profundo permiten optimizar procesos, analizar grandes volúmenes de datos, automatizar tareas y personalizar soluciones. En particular, algoritmos como SVM, redes neuronales, K-means y Deep Learning han mostrado ser efectivos en tareas de predicción, clasificación y agrupamiento, lo que evidencia su impacto en la mejora de la eficiencia y precisión en distintos ámbitos del desarrollo de software.

¿Cuáles son las limitaciones para la adopción del aprendizaje automático en el proceso de desarrollo de software?

Las principales limitaciones para la adopción del aprendizaje automático en el proceso de desarrollo de software están relacionadas con la calidad de los datos, ya que requieren grandes volúmenes de datos etiquetados, lo cual puede ser costoso y difícil de obtener, la alta demanda de recursos computacionales es otra limitante ya que para entrenar y ejecutar modelos complejos puede ser una barrera significativa, especialmente en entornos con recursos limitados, otro obstáculo es el ajuste de hiperparámetros, un proceso crucial pero complicado que requiere considerable experimentación y conocimiento especializado.

La interpretabilidad y transparencia de los modelos de AA son otras limitaciones importantes, especialmente con los modelos de aprendizaje profundo considerados "cajas negras," lo que dificulta la comprensión y justificación de las decisiones tomadas por estos modelos. Además, la capacidad de los modelos para generalizar bien a datos nuevos y no vistos es un problema común, con el sobreajuste siendo una preocupación frecuente.

El sesgo en los datos y modelos puede llevar a resultados injustos o discriminatorios, planteando problemas de equidad. A pesar de estos desafíos, el potencial transformador del AA en el desarrollo de software es innegable, pero es crucial abordar estas limitaciones para maximizar su impacto y efectividad. En la siguiente tabla se visualiza cada autor con su limitante que ellos trataron.

Tabla 5 Limitaciones para aplicar el Aprendizaje Automático

N.º	Limitación	Referencia
1	Calidad y cantidad de datos	(Valdivieso et al., 2019) (Para et al., 2020) (Cadena Lema, 2020) (Rojas, n.d., 2020) (Gante Diaz et al., 2022) (Mejía et al., 2023)
2	Recursos computacionales	(Bazzana Tanevitch et al., 2021) (Rojas, n.d., 2020) (Chuquimarca Jimenez et al., 2021)
3	Ajuste de hiperparámetros	(Contreras Contreras et al., 2022) (Bazzana Tanevitch et al., 2021) (Flórez-Martíne et al., 2020)
4	Interpretabilidad y transparencia	(Valdivieso et al., 2019) (Polo-Triana et al., 2022) (Patiño et al., 2023) (Chuquimarca Jimenez et al., 2021)
5	Generalización y sobreajustes	(Bazzana Tanevitch et al., 2021)
6	Sesgo en los datos y modelos	(Omar et al., 2022) (Gil M. n.d., 2023)

En conclusión, la adopción del aprendizaje automático en el desarrollo de software enfrenta varias limitaciones clave, como la necesidad de grandes volúmenes de datos etiquetados, la alta demanda de recursos computacionales, y la complejidad del ajuste de hiperparámetros. Además, la falta de interpretabilidad en algunos modelos, especialmente en el aprendizaje profundo, y los

problemas de generalización y sesgo en los datos presentan barreras significativas. Superar estos desafíos es fundamental para que el aprendizaje automático pueda alcanzar su máximo potencial y brindar soluciones más justas, eficientes y transparentes en el desarrollo de software.

Discusión

El aprendizaje automático ha transformado el desarrollo de software, facilitando la automatización de tareas complejas, mejorando la precisión y eficiencia en la toma de decisiones, y proporcionando nuevas capacidades analíticas. Ha permitido el desarrollo de sistemas más inteligentes y adaptativos, optimizando procesos en diversas industrias, desde la salud y la agricultura hasta el comercio electrónico y la educación. Sin embargo, la implementación de estas técnicas también enfrenta desafíos significativos, como la dependencia de la calidad de los datos, la necesidad de recursos computacionales elevados y la complejidad en el ajuste de modelos y algoritmos. La adopción de técnicas de aprendizaje automático continúa creciendo, impulsada por la creciente disponibilidad de datos y avances en la tecnología, pero es esencial abordar estas limitaciones para maximizar su potencial en el desarrollo de software.

Para futuras investigaciones, es esencial mejorar la gestión y calidad de los datos, optimizar los recursos computacionales, y desarrollar modelos más explicables y transparentes. La automatización en la selección y ajuste de modelos mediante aprendizaje automático automatizado, junto con la seguridad y privacidad, son áreas críticas a explorar. Integrar tecnologías emergentes como IoT y blockchain, así como también, se deben evaluar las implicaciones éticas y sociales del aprendizaje automático, también son esenciales para maximizar su potencial y asegurar un impacto positivo en el desarrollo de software.

CONCLUSIONES

- El aprendizaje automático ha transformado significativamente el desarrollo de software, facilitando la automatización de tareas complejas y mejorando la precisión y eficiencia en la toma de decisiones. Esta tecnología ha permitido el desarrollo de sistemas más inteligentes y adaptativos, optimizando procesos en diversas industrias, como, por ejemplo, en el sector de la salud, la agricultura, comercio electrónico, educación, entre otros.
- La implementación de estas técnicas enfrenta desafíos significativos. La calidad de los datos es crucial para el éxito de los modelos, y la dependencia de datos de alta calidad puede ser una limitación. Además, los recursos computacionales elevados necesarios para entrenar modelos complejos y la dificultad en el ajuste de estos modelos y algoritmos son barreras importantes. A pesar de estos desafíos, la adopción de técnicas de aprendizaje automático en el desarrollo de software continúa creciendo, impulsada por la creciente disponibilidad de datos y avances en la tecnología.
- El futuro del desarrollo de software con aprendizaje automático es prometedor. Las investigaciones continuas y los desarrollos tecnológicos seguirán mejorando las capacidades de aprendizaje automático, permitiendo la creación de software aún más inteligente y eficiente. La colaboración entre investigadores y profesionales del software será crucial para superar los desafíos actuales y aprovechar al máximo las oportunidades que ofrece esta tecnología.
- Considerar las implicaciones éticas y sociales de la aplicación del aprendizaje automático en el desarrollo de software. La transparencia, la equidad y la responsabilidad deben ser principios clave en el diseño y despliegue de sistemas basados en aprendizaje automático, para garantizar que los beneficios se distribuyan equitativamente y se minimicen los riesgos.
- El aprendizaje automático ha revolucionado el desarrollo de software, ofreciendo enormes beneficios y oportunidades, pero también presentando desafíos que deben ser abordados. Con un enfoque equilibrado en la investigación, la innovación y la ética, el aprendizaje automático continuará desempeñando un papel crucial en el avance del desarrollo de software y en la optimización de procesos en diversas industrias.

RECOMENDACIONES

- Se recomienda la utilización de la metodología PRISMA para asegurar una revisión exhaustiva y sistemática de la literatura, nos proporciona una base sólida para futuras investigaciones y esta metodología nos ayudará a identificar de manera precisa las áreas de estudio que requieren más atención y a sintetizar los hallazgos de manera coherente.
- Debido a la precisión y efectividad de los modelos de aprendizaje automático dependen en gran medida de la calidad de los datos, es crucial implementar estrategias robustas para la recopilación, limpieza y preprocesamiento de datos, las investigaciones futuras deben enfatizar el desarrollo de métodos para el manejo de datos atípicos y hacerlo más legible y coherente.
- Los modelos de aprendizaje automático, especialmente los que emplean redes neuronales profundas, requieren recursos computacionales significativos, para esto se recomienda explorar métodos para reducir la complejidad computacional, como técnicas de compresión de modelos y optimización de algoritmos, para facilitar la implementación en entornos con recursos limitados.
- Para abordar las limitaciones y desafíos identificados, es beneficioso enfocar futuras investigaciones en aplicaciones específicas del aprendizaje automático que puedan proporcionar soluciones prácticas a problemas concretos, esto podría incluir la personalización de técnicas para dominios particulares como la salud, la agricultura, o el comercio electrónico.

BIBLIOGRAFÍA

Arias, E. R. (2024, March 19). Investigación descriptiva: Qué es, tipos y ejemplos. Economipedia. <https://economipedia.com/definiciones/investigacion-descriptiva.html>

Agama Espinoza, A. S. (2021). Revisión de la literatura del comercio electrónico, el aprendizaje automático y sus aplicaciones en la industria y tiendas por departamento en línea.

<https://tesis.pucp.edu.pe/repositorio//handle/20.500.12404/20714>

Amazon.com. Recuperado el 18 de mayo de 2024, (S/f). de <https://aws.amazon.com/es/what-is/cybersecurity/#:~:text=La%20ciberseguridad%20es%20la%20pr%C3%A1ctica,cliente%20y%20cumplir%20la%20normativa>.

Bazzana Tanevitch, L., Dioguardi, F., Delle Ville, J., Villena, S., Herrera, F., Hasperué, W., Torres, D., & Antonelli, R. L. (2021). Aprendizaje automático aplicado al procesamiento de imágenes para la clasificación de objetos reciclables. *Memorias Del Congreso Argentino En Ciencias de La Computación - CACIC 2021*, 1(1), 31–40. <http://sedici.unlp.edu.ar/handle/10915/130312>

Burns, E. (2021, septiembre 6). Aprendizaje profundo (deep learning). ComputerWeekly.es; TechTarget. <https://www.computerweekly.com/es/definicion/Aprendizaje-profundo-deep-learning>

Cadena Lema, H. D. (2020). Diseño de un sistema para el control de riego mediante técnicas de aprendizaje automático aplicada a la agricultura de precisión en

la granja La Pradera de la universidad Técnica del Norte.

<https://repositorio.utn.edu.ec/handle/123456789/10296>

Cedeno-Moreno, D., & Vargas-Lombardo, M. (2020). Aprendizaje automático aplicado al análisis de sentimientos. *I+D Tecnológico*, 16(2), 59–66.

<https://doi.org/10.33412/IDT.V16.2.2833>

Chuquimarca Jimenez, L. E., Pinzón Tituana, S., & Rosales Pincay, A. (2021). Detección de mascarilla para COVID-19 a través de Aprendizaje Profundo usando OpenCV y Cascade Trainer GUI. *Revista Científica y Tecnológica UPSE*, 8(1), 68–73.

<https://doi.org/10.26423/RCTU.V8I1.572>

Contreras Contreras, G. F., Medina Delgado, B., Acevedo Jaimes, B. R., & Guevara Ibarra, D. (2022). Metodología de desarrollo de técnicas de agrupamiento de datos usando aprendizaje automático. *Tecnura: Tecnología y Cultura Afirmando El Conocimiento*, ISSN-e 2248-7638, ISSN 0123-921X, Vol. 26, No. 72, 2022, Págs. 42-58, 26(72), 42–58. <https://doi.org/10.14483/22487638.17246>

Di Pierro, G. (2024, January 3). La evolución del desarrollo de software en los últimos 20 años - Diciembre. <https://es.linkedin.com/pulse/la-evoluci%C3%B3n-del-desarrollo-de-software-en-los-20-a%C3%B1os-di-pierro-3ulbf>

Del Ejército esquina, A. 9. de O. N. 1616 y. A., & del Guayas, E. C. del C. de A. (s/f). DE LA ECONOMÍA. Wipo.int. Recuperado el 20 de mayo de 2024, de <https://www.wipo.int/edocs/lexdocs/laws/es/ec/ec075es.pdf>

De, L. 0. R. O. S. 899. (s/f). CODIGO ORGANICO DE LA ECONOMIA SOCIAL DE LOS CONOCIMIENTOS. Gob.ec. Recuperado el 20 de mayo de 2024,

de <https://www.ambiente.gob.ec/wp-content/uploads/downloads/2018/05/Codigo-Organico-Economia-Social-de-los-Conosimientos.pdf>

De Ceupe, B. (2022, March 28). Ceupe. Ceupe.

<https://www.ceupe.com/blog/aprendizaje-supervisado.html?dt=1716160599247>

Enrique, V. L. G. (2022). Utilización de Machine Learning para el proceso de selección de personal en una microempresa.

<http://dspace.ups.edu.ec/handle/123456789/23335>

Flórez-Martíne, A., Vargas-Flórez, J. O., Perez-Waltero, H. E., & Quintana-Fuentes, L. F. (2020). Análisis de componentes principales utilizando python para identificar clúster asociados a muestras de cacao seco sano e infectado con monilia en Norte de Santander. *AiBi Revista de Investigación, Administración e Ingeniería*, 8(2), 16–22. <https://doi.org/10.15649/2346030X.712>

Fernández, E. C. (2024, 26 febrero). Conoce la historia del machine learning: ¡desde sus inicios! Tokio School. <https://www.tokioschool.com/noticias/historia-machine-learning/#historia-del-machine-learning-comienzos-en-el-desarrollo-del-aprendizaje-automatico>

Fernández, E. C. (2024, February 26). Conoce la historia del machine learning: ¡desde sus inicios! Tokio School. <https://www.tokioschool.com/noticias/historia-machine-learning/>

Gil-Vera, V. D., & Seguro-Gallego, C. (2022). Machine learning aplicado al análisis del rendimiento de desarrollos de software. *Revista Politécnica*, 18(35), 128-139. <https://doi.org/10.33571/rpolitec.v18n35a9>

GAMCO, SL. Recuperado el 20 de mayo de 2024. Qué es Aprendizaje semisupervisado Concepto y definición. Glosario. (s/f)., de <https://gamco.es/glosario/aprendizaje-semisupervisado/>

Henríquez Miranda, C., Salcedo Morillo, D. D., & Sánchez Torres, G. (2022). El aprendizaje automático en entornos educativos universitarios: caso deserción académica. *Prospectiva*, ISSN-e 2216-1368, ISSN 1692-8261, Vol. 20, No. 1, 2022, 20(1), 4. <https://doi.org/10.15665/rp.v20i1.2736>

ISO - Aprendizaje automático (AA): todo lo que hay que saber. (n.d.). Retrieved September 15, 2024, from <https://www.iso.org/es/inteligencia-artificial/aprendizaje-autom%C3%A1tico>

ISO. (2024). Aprendizaje automático: todo lo que hay que saber. <https://www.iso.org/es/contents/news/insights/AI/machine-learning-need-to-know.html>

Mejía, J. A., Oviedo-Benalcázar, M. A., Ordoñez, J. A., & Valencia, J. F. (2023). Machine learning applied to the prediction of diabetes mellitus, using socioeconomic and environmental information from health system users. *Revista Facultad Nacional de Salud Publica*, 41(2). <https://doi.org/10.17533/UDEA.RFNSP.E351168>

Método PRISMA: qué es y cómo utilizarlo - Tesis Doctorales Online. (n.d.). Retrieved July 26, 2024, from <https://tesisdoctoralesonline.com/metodo-prisma-que-es-y-como-usarlo-en-una-revision-sistemica/>

Medina, O., & Maybee, A. (2018). Detección de depresión a través de análisis textual utilizando aprendizaje automático, 2017. Universidad Ricardo Palma.

<https://hdl.handle.net/20.500.14138/1674>

Navarro, S. (2022, agosto 29). Aprendizaje supervisado y no supervisado en data analytics. KeepCoding Bootcamps. <https://keepcoding.io/blog/aprendizaje-supervisado-y-no-supervisado/>

Natalia. (2022, May 27). ¿Qué es el aprendizaje por refuerzo? <https://www.ceupe.cl/blog/que-es-el-aprendizaje-por-refuerzo.html>

Norman, A. T. (2019). Aprendizaje Automático En Acción. Litres.

Omar, M., Germán, V., & Dima, C. (2022). Desarrollo de una herramienta de aprendizaje automático (machine learning) para establecer relaciones entre ocupaciones y programas de capacitación en el Uruguay. CEPAL.

<https://hdl.handle.net/11362/47724>

Oracle Argentina. (n.d.). Retrieved July 28, 2024, from ¿Qué es el aprendizaje automático? <https://www.oracle.com/ar/artificial-intelligence/machine-learning/what-is-machine-learning/>

Ojeda, S., & Isabel, M. (2021). Algoritmos de aprendizaje automático para predecir la curación de malformaciones arteriovenosas cerebrales tratadas con radiocirugía. Universidad Ricardo Palma.

Oracle Argentina. (n.d.). Retrieved July 28, 2024, from ¿Qué es el aprendizaje automático? <https://www.oracle.com/ar/artificial-intelligence/machine-learning/what-is-machine-learning/>

Para, T., El, O., & De, G. (2020). Recocido simulado y aprendizaje automático aplicado a la optimización de factores de riesgo empresarial.

<http://riaa.uaem.mx/xmlui/handle/20.500.12055/1243>

Patiño, D., Medina, J., Silva, R., Guijarro, A., & Rodríguez, J. (2023).

PREDICTION OF ARRHYTHMIAS AND ACUTE MYOCARDIAL

INFARCTIONS USING MACHINE LEARNING. *Ingenius*, 2023(29), 79–89.

<https://doi.org/10.17163/INGS.N29.2023.07>

Polo-Triana, S. I., Ramírez-Sierra, Y. A., Arias-Osorio, J. E., Martínez-Vega, R. A., & Lamos-Díaz, H. (2022). Métodos de aprendizaje automático para predecir el comportamiento epidemiológico de enfermedades arbovirales: revisión estructurada de literatura. *Salud UIS*, 55(1). <https://doi.org/10.18273/SALUDUIS.55.E:23017>

Paico Chileno, D., & Valdera Contreras, J. H. (2021). Detección automática de ataques de inyección mediante aprendizaje automático en bases de datos NOSQL.

Universidad Señor de Sipán. <https://hdl.handle.net/20.500.12802/8908>

Rojas, E. M. (n.d.). Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo.

Valdivieso, A., Díaz, C., & Sarmiento, J. (2019). Mhealth con bigdata y machine learning como soporte tecnológico para la salud en Colombia. *Loginn*, ISSN-e 2590-7441, Vol. 3, No. 2, 2019, 3(2), 6.

<https://dialnet.unirioja.es/servlet/articulo?codigo=8738582&info=resumen&idioma=ENG>

Torres, B. (2019, July 24). ¿Qué es el aprendizaje automático y cómo funciona? - UNAM Global. UNAM Global - De La Comunidad Para La Comunidad.

https://unamglobal.unam.mx/global_revista/que-es-el-aprendizaje-automatico-y-como-funciona/#:~:text=El%20aprendizaje%20autom%C3%A1tico%20no%20es,a%20la%20de%20los%20humanos.

UNIE. (2023, 28 noviembre). ¿Cuál es la verdadera diferencia entre Inteligencia Artificial y Machine Learning? UNIE.

<https://www.universidadunie.com/blog/diferencia-ia-y-machine-learning>

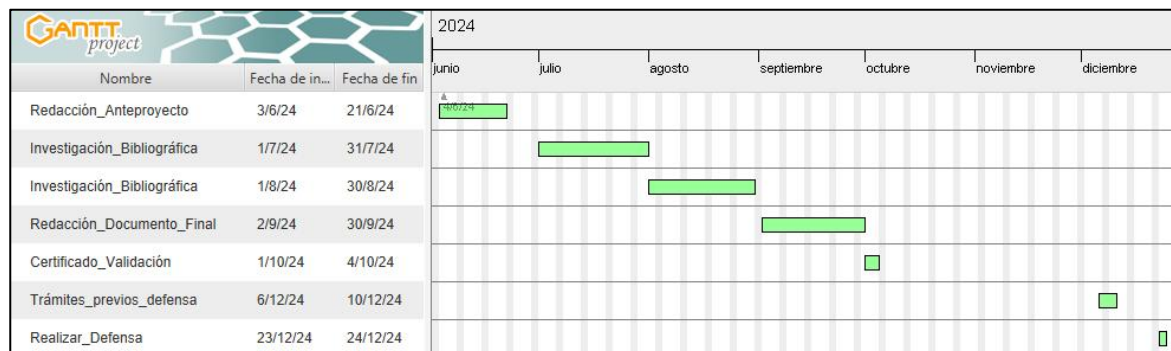
Visor Redalyc - Predicción de la presión de burbujeo utilizando aprendizaje automático. (n.d.). Retrieved July 18, 2024, from

<https://www.redalyc.org/journal/6738/673874721014/>

ANEXOS

Cronograma (Gantt)

Figura 3 Cronograma de Gantt



Presupuesto Ejecutado

Tabla 6 Presupuesto

N	Recursos	Cantidad	Precio Unitario	Total
1	Computador Core I5 8GB RAM	1	300	300
2	Laptop	1	300	300
3	Impresora	1	350	350
4	Internet	5	25	125
5	Pasajes	15	2.85	42.75
6	Energía eléctrica	5	15	75
Total				1.192,75

- Certificado Antiplagio



FACULTAD DE CIENCIAS
ADMINISTRATIVAS,
GESTIÓN EMPRESARIAL
E INFORMÁTICA

**ING. EDGAR RIVADENEIRA RAMOS EN CALIDAD DE DIRECTOR
DEL TRABAJO DE INTEGRACIÓN CURRICULAR,**

CERTIFICA

Que el trabajo de integración curricular denominado "EL APRENDIZAJE AUTOMÁTICO APLICADO AL DESARROLLO DE SOFTWARE.", presentado por RICARDO JAVIER CHORA PAUCAR estudiante de la **carrera de Software** pasó el análisis de coincidencia no accidental en la herramienta TURNITING, reflejando un **porcentaje de similitud del 3%**, como se puede evidenciar en el documento adjunto.

Guaranda, 09 de diciembre del 2024

Atentamente,



Ing. Edgar Rivadeneira
Director

TESIS_FINAL_DEFENSA_RICARDO_1.docx

 Universidad Estatal de Bolívar

Detalles del documento

Identificador de la entrega

trcaid:3117414279429

54 Páginas

Fecha de entrega

9 dic 2024, 11:45 a.m. GMT-5

10,255 Palabras

Fecha de descarga

9 dic 2024, 12:59 p.m. GMT-5

65,130 Caracteres

Nombre de archivo

TESIS_FINAL_DEFENSA_RICARDO_1.docx

Tamaño de archivo

3.5 MB

3% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes suprimidas, para co...



Filtrado desde el informe

- Bibliografía
- Todo el texto
- Texto mencionado
- Coincidencias menores (menos de 12 palabras)

Exclusiones

- N.º de fuentes excluidas
- N.º de coincidencias excluidas

Fuentes principales

- 2%  Fuentes de Internet
- 0%  Publicaciones
- 3%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para hacer coincidencias que permitirán distinguirlo de uno original. Si además algo resulta, lo marcamos con una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y lo revise.



Edgar Brodensen Pangu