



Sistema automatizado de reportes académicos que optimiza la gestión educativa

Automated Academic Reporting System that Optimizes Educational Management

Jennifer Andrea Londoño-Gallego, Esp. en Formulación y evaluación de proyectos, Servicio Nacional de Aprendizaje SENA, Medellín (Colombia) (jealondonog@sena.edu.co) (<https://orcid.org/0000-0003-2957-9178>)
Daniel David Benavides-Sánchez, Mag. en Gestión de Tecnologías de la Información, Servicio Nacional de Aprendizaje SENA, Medellín (Colombia) (dbenavidess@sena.edu.co) (<https://orcid.org/0000-0002-5202-8537>)
John Jairo Castro-Maldonado*, Dr. en Educación, Servicio Nacional de Aprendizaje SENA, Medellín (Colombia) (jcastrom@sena.edu.co) (<https://orcid.org/0000-0002-3823-4297>)
Oscar Eduardo García-Quintero, Mag. en Ciencias en la especialidad de Matemáticas, Servicio Nacional de Aprendizaje SENA, Medellín (Colombia) (oegarcia@sena.edu.co) (<https://orcid.org/0009-0002-4572-366X>)

* Indicates the corresponding author

RESUMEN

En el contexto de transformación digital, las instituciones de formación profesional enfrentan vacíos en la gestión educativa relacionados con la programación, el seguimiento académico y la generación de reportes para la toma de decisiones. Este artículo presenta el Sistema Automatizado de Reportes Académicos (SARA), cuyo objetivo es describir su desarrollo técnico e implementación de inteligencia artificial, así como evaluar su aporte a la gestión académica mediante análisis bibliométrico y la percepción de instructores. La investigación es de tipo aplicada, con enfoque mixto y alcance descriptivo-exploratorio. El marco arquitectónico de SARA se fundamenta en el patrón Modelo-Vista-Controlador (MVC) y en una base de datos NoSQL, lo que asegura escalabilidad modular y la integración eficiente de más de 140.000 registros heterogéneos (Excel, SOFIA Plus, JSON). Para el análisis bibliométrico se consultó la base de datos Scopus, identificando 4.922 documentos procesados con VOSviewer® aplicando análisis de coautoría. Adicionalmente, se aplicaron entrevistas semiestructuradas a 7 instructores, cuyo análisis se apoyó en técnicas de procesamiento de lenguaje natural para sistematizar percepciones. Los resultados destacan en la literatura la relevancia de marcos de interoperabilidad, gobernanza de datos y sostenibilidad digital, además de evidenciar una percepción positiva respecto a la facilidad de uso, la reducción de carga administrativa y la utilidad de SARA en la toma de decisiones. En conclusión, SARA se plantea como una herramienta escalable y replicable que fortalece la gestión académica y genera implicaciones pedagógicas y organizativas desde un enfoque interdisciplinario.

ABSTRACT

In the context of digital transformation, vocational education institutions face gaps in academic management related to scheduling, learner monitoring, and the generation of reports for decision-making. This article introduces the Automated Academic Reporting System (SARA), whose objective is to describe its technical development and implementation of artificial intelligence, as well as to assess its contribution to academic management through bibliometric analysis and instructor perceptions. The research is applied in nature, with a mixed-methods approach and a descriptive-exploratory scope. SARA's architectural framework is based on the Model-View-Controller (MVC) pattern and a NoSQL database, ensuring modular scalability and the efficient integration of more than 140,000 heterogeneous records (Excel, SOFIA Plus, JSON). For the bibliometric analysis, the Scopus database was consulted, identifying 4,922 documents processed with VOSviewer® using co-authorship analysis. In addition, semistructured interviews were conducted with seven instructors, whose responses were analyzed using natural language processing techniques to systematize perceptions. The findings highlight in the literature the relevance of frameworks for interoperability, data governance, and digital sustainability, while also revealing a positive perception of SARA regarding ease of use, reduction of administrative workload, and utility in decision-making. In conclusion, SARA is presented as a scalable and replicable tool that strengthens academic management and generates pedagogical and organizational implications from an interdisciplinary perspective.

PALABRAS CLAVE | KEYWORDS

Gestión educativa, sistema automatizado, experiencia de usuario, análisis de clúster, inteligencia artificial.
Educational Management, Automated System, User Experience, Cluster Analysis, Artificial Intelligence.

1. Introducción

La optimización de la gestión educativa se entiende como la capacidad de una institución para coordinar, integrar y mejorar de manera continua los procesos académicos y administrativos, con el fin de garantizar eficiencia operativa, calidad pedagógica y pertinencia en los resultados de aprendizaje. En un contexto de transformación digital, este concepto trasciende la mera digitalización de trámites y apunta a la automatización inteligente, la gobernanza de datos y la toma de decisiones basada en evidencia (Mendieta Lucas et al., 2025).

La literatura reciente enfatiza que los principales retos de la gestión académica en instituciones de formación profesional se relacionan con la fragmentación de datos, la dependencia de procesos manuales y la limitada interoperabilidad entre sistemas (García-Peñalvo y Corell, 2020; Holmes et al., 2019). Estas limitaciones reducen la capacidad de respuesta de las instituciones frente a fenómenos como la deserción estudiantil, la planificación curricular o la trazabilidad de aprendizajes. De acuerdo con la OECD (2019) y la UNESCO (2022), la optimización de la gestión educativa se puede lograr a través de tres estrategias principales: la automatización de procesos administrativos y académicos para reducir tiempos y errores; la consolidación de información académica en repositorios interoperables que faciliten la trazabilidad; y la incorporación de sistemas de análisis predictivo y reportes en tiempo real que permitan anticipar decisiones estratégicas. En este sentido, diversos estudios empíricos han demostrado que la implementación de soluciones basadas en arquitecturas escalables y orientadas a datos contribuye a mejorar la gestión institucional. Por ejemplo, el uso de sistemas de analítica de aprendizaje y minería de datos educativos ha permitido detectar patrones de desempeño, prevenir la deserción y personalizar apoyos pedagógicos (Camones Gonzales et al., 2024; Romero y Ventura, 2020).

En Latinoamérica, investigaciones recientes evidencian la necesidad de sistemas integrados que reduzcan la dependencia de plataformas fragmentadas y aumenten la capacidad de los docentes y coordinadores para acceder a información consolidada (Kung et al., 2024). La transición hacia entornos educativos más digitalizados no solo implica una actualización de herramientas, sino también la redefinición de procesos esenciales como la recolección, el análisis y la presentación de datos académicos (Arias Ortiz et al., 2021). El crecimiento exponencial en el volumen de datos que manejan actualmente las empresas y en especial las instituciones educativas de formación profesional integral, exige transformar dicha información en conocimiento útil, veraz y pertinente (Mustafa et al., 2024).

En instituciones educativas, la gestión de la información académica va más allá del simple registro de asistencias y novedades, incluye variables como el desempeño disciplinario y el cumplimiento de objetivos de aprendizaje. Este enfoque permite una comprensión más profunda de las trayectorias estudiantiles, facilitando la identificación de necesidades y la personalización del acompañamiento formativo. Es así como la optimización de estos procesos en las instituciones educativas se ha vuelto una necesidad imperativa para asegurar una experiencia formativa de alta calidad y una gestión de recursos eficiente.

En este contexto, los sistemas automatizados de información adquieren relevancia al ofrecer marcos estructurados que permiten abordar tareas rutinarias y operativas, reduciendo tiempos de ejecución, mejorando la trazabilidad, fomentando la accesibilidad y optimizando la gestión de procesos (Nelson Salgado, 2024). El Servicio Nacional de Aprendizaje (SENA), como caso puntual del presente estudio, es una institución pública en Colombia dedicada a la formación profesional integral que ofrece formación gratuita en programas técnicos, tecnológicos y complementarios orientados al desarrollo de competencias laborales y al fortalecimiento del talento humano en distintos sectores productivos, su operación se articula a través de Centros de Formación distribuidos en distintas regiones, apoyados por plataformas como *SOFIA PLUS* - Sistema Optimizado para la Formación y el Aprendizaje Activo, que ayuda en la administración, gestión, operación y evaluación del proceso educativo, desde el registro del aprendiz hasta su certificación, *Zajuna*, otra plataforma de aprendizaje virtual que facilita la adaptación de la comunidad educativa a la formación virtual frente al seguimiento de actividades concertadas entre aprendices e instructores y el *SGVA* sistema virtual de aprendices cuyo fin es gestionar todos los procesos de formación, desde la inscripción hasta el seguimiento académico de los aprendices, incluyendo la gestión de contratos de aprendizaje.

Aunque estas herramientas son fundamentales, existen vacíos significativos en la gestión de otros procesos estructurales no cubiertos por dichas plataformas, lo cual limita la capacidad operativa de cada centro de formación. Particularmente, la imposibilidad de gestionar procesos internos específicos como la programación académica, el seguimiento detallado del rendimiento académico de los aprendices o la asignación de instructores. Esta falta de capacidad para gestionar internamente obstaculiza la eficiencia

operativa y la calidad de la experiencia educativa ofrecida por cada centro de formación del SENA, lo que evidencia la necesidad de una solución integral, flexible y escalable.

En respuesta a esta problemática, se desarrolla el Sistema Automatizado de Reportes Académicos (SARA), una plataforma digital diseñada para optimizar la gestión de información académica. SARA integra un conjunto de procesos formales sobre una base de datos estructurada que permite generar, analizar y distribuir información crítica para la operación institucional y la toma de decisiones informadas. Además, incorpora elementos de analítica de datos e inteligencia artificial, lo que fortalece la eficiencia, la eficacia y la productividad del proceso formativo.

La justificación de este trabajo se centra en que, a pesar de los esfuerzos previos, en instituciones como el SENA aún persisten vacíos operativos que obstaculizan una gestión eficiente: coexistencia de múltiples fuentes de datos (Excel, SOFIA Plus, JSON), falta de integración entre sistemas y limitaciones en la generación de reportes oportunos. Estas condiciones refuerzan la pertinencia de desarrollar modelos como SARA, que combinan arquitectura modular (MVC + NoSQL) con capacidades de consolidación e interoperabilidad. La elección de este tema responde a tres motivaciones principales. En primer lugar, los vacíos persistentes en la gestión académica de instituciones de formación profesional, que dificultan la programación, el seguimiento de aprendices y la generación de reportes confiables. En segundo lugar, la oportunidad de articular innovación tecnológica con políticas educativas, dado que organismos internacionales como la UNESCO y la OCDE recomiendan fortalecer la gobernanza de datos y la toma de decisiones basadas en evidencia. Finalmente, la necesidad de contar con modelos replicables y escalables, que puedan ser adoptados por distintos centros del SENA y otras instituciones de formación.

Bajo esta perspectiva, el propósito del estudio es doble, primeramente, en el desarrollo técnico del sistema SARA, donde se detalla su arquitectura e implementación de inteligencia artificial y se justifica el aporte en la gestión educativa con estudios teóricos, analizados de forma bibliométrica teniendo en cuenta recomendaciones de política institucional que permitan su sostenibilidad y expansión. Por otro lado, la percepción de la herramienta por parte de los instructores SENA en sus actividades académicas, cuyos resultados son procesados a través de técnicas de análisis de sentimientos. En conjunto, se busca aportar al cuerpo de conocimientos sobre la automatización de procesos y transformación digital en la educación, ofreciendo una solución replicable y adaptativa para otras instituciones de educación superior interesadas en la optimización de sus procesos de gestión educativa.

Se espera, además, que herramientas como SARA no solo respondan a necesidades puntuales, sino que se integren con los demás sistemas existentes en la institución, operando de manera unificada para generar impactos más amplios y diferenciados dentro de la comunidad educativa, impactos que incluyen la mejora de la trazabilidad académica, una mayor capacidad para prevenir la deserción formativa, la optimización de recursos humanos y físicos, y el fortalecimiento de la planificación pedagógica estratégica, todo ello orientado al logro de una educación más pertinente, inclusiva y sostenible.

En suma, este estudio contribuye principalmente al campo de la educación y la gestión pedagógica, al presentar y evaluar un sistema que apoya la programación, el seguimiento y la toma de decisiones en un centro de formación. Su valor añadido radica en su carácter interdisciplinario, que combina marcos de la ingeniería de software y técnicas de procesamiento de lenguaje natural con un enfoque pedagógico orientado a la calidad educativa. Así, el artículo aporta evidencia aplicada sobre cómo la convergencia entre tecnología y educación puede generar impactos significativos en la gestión académica y en la formulación de políticas institucionales.

1.1. Desarrollo del sistema automatizado de reportes académicos (SARA)

Un sistema automatizado de información se compone de dos tipos principales de elementos: los elementos técnicos y los elementos funcionales. Los primeros incluyen las entidades físicas y lógicas que conforman el sistema, tales como dispositivos computacionales, software, personal técnico y bases de datos. Los segundos, por su parte, comprenden las funciones operativas del sistema, que abarcan desde el procesamiento de transacciones y la administración de datos, hasta la generación y presentación de información útil para la toma de decisiones (Rivera Aguilera et al., 2024).

La construcción del sistema SARA partió de un análisis contextual que identificó problemáticas asociadas a la programación de instructores y al seguimiento académico de aprendices. A partir de este diagnóstico, se definieron los requerimientos técnicos y funcionales del sistema, seguidos por el diseño, la

implementación y la posterior validación operativa de la herramienta. Inicialmente, la alimentación del sistema se basó en datos extraídos de múltiples hojas de cálculo (Excel), centradas en la programación académica. Posteriormente, se incorporaron reportes estructurados provenientes de la plataforma institucional Sena Sofia Plus, los cuales contienen información crítica sobre el rendimiento académico y los datos personales de los aprendices. Se estima que estos reportes generan más de 140.000 registros, los cuales deben ser procesados y almacenados de manera eficiente dentro de la infraestructura del sistema.

Debido a la complejidad del volumen de datos y la necesidad de procesamiento eficiente, el sistema fue desarrollado utilizando el lenguaje de programación Python, en combinación con la biblioteca Pandas, ampliamente reconocida por su robustez en la manipulación y análisis de datos estructurados. Para el almacenamiento, se optó por una base de datos NoSQL, dada su capacidad para manejar grandes volúmenes de datos no estructurados y semi-estructurados, así como su flexibilidad para adaptarse a cambios en los esquemas de datos durante el crecimiento y evolución del proyecto. Este enfoque tecnológico no solo garantiza eficiencia en el procesamiento, sino que también permite escalar el sistema y adaptarlo a diferentes necesidades institucionales, lo cual resulta fundamental en entornos educativos en constante transformación. SARA, en este sentido, representa una solución adaptativa y dinámica para la gestión académica, orientada por principios de automatización, analítica avanzada y toma de decisiones basada en evidencia. (Figura 1)

A continuación, se describen algunos fragmentos del proceso de codificación usado en el sistema.

Figura 1: Fragmento de código del sistema SARA.

```
df = pd.read_excel(f'static/horarios/{trimestre}.xlsx')
df = df.drop(0)
df.columns = ['source', 'FICHA', 'FORMACION', 'TITULAR', 'TRIMESTRE', 'COMPETENCIA', 'NOMBRE_COMPETENCIA', 'RAP_1', 'RAP_2', 'RAP_3', 'RAP_4', 'RAP_5', 'RAP_6',
'INSTRUCTOR', 'IC2', 'HORAS_SEMANAL', 'IB2', 'h6_71', 'h7_81', 'h8_91', 'h9_101', 'h10_111', 'h11_121', 'h12_131', 'h13_141', 'h14_151', 'h15_161', 'h16_171',
'h17_181', 'h18_191', 'h19_201', 'h20_211', 'h21_221', 'h6_79', 'h7_89', 'h8_99', 'h9_109', 'h10_119', 'h11_129', 'h12_139', 'h13_149', 'h14_159', 'h15_169',
'h16_179', 'h17_189', 'h18_199', 'h19_209', 'h20_219', 'h21_229', 'h6_791', 'h7_891', 'h8_991', 'h9_1091', 'h10_1191', 'h11_1291', 'h12_1391', 'h13_1491',
'h14_1591', 'h15_1691', 'h16_1791', 'h17_1891', 'h18_1991', 'h19_2091', 'h20_2191', 'h21_2291', 'h6_73', 'h7_83', 'h8_93', 'h9_103', 'h10_113', 'h11_123',
'h12_133', 'h13_143', 'h14_153', 'h15_163', 'h16_173', 'h17_183', 'h18_193', 'h19_203', 'h20_213', 'h21_223', 'h6_79', 'h7_89', 'h8_99', 'h9_109', 'h10_119', 'h11_129',
'h12_139', 'h13_149', 'h14_159', 'h15_169', 'h16_179', 'h17_189', 'h18_199', 'h19_209', 'h20_219', 'h21_229', 'h6_75', 'h7_85', 'h8_95', 'h9_105',
'h10_115', 'h11_125', 'h12_135', 'h13_145', 'h14_155', 'h15_165', 'h16_175', 'h17_185', 'h18_195', 'h19_205', 'h20_215', 'h21_225']
df = df.drop(['source', 'TITULAR', 'COMPETENCIA', 'RAP_1', 'RAP_2', 'RAP_3', 'RAP_4', 'RAP_5', 'RAP_6', 'IC2', 'HORAS_SEMANAL', 'IB2', 'HORAS_SEMANAL'], axis=1)
df['NOMBRE_COMPETENCIA'] = df['NOMBRE_COMPETENCIA'].str.split('-').str[-1].str.strip()
```

Fuente propia

El código anterior realiza la lectura de un archivo de Excel que contiene la programación de instructores y aprendices, cargándolo en un DataFrame de Pandas. Se lleva a cabo un proceso de ETL (Extracción, Transformación y Limpieza de Datos), donde se eliminan algunas filas y columnas innecesarias. El algoritmo da para que el sistema unificado sea capaz de transportar la información a escenarios educativos permitiendo que el sistema unificado procese la data.

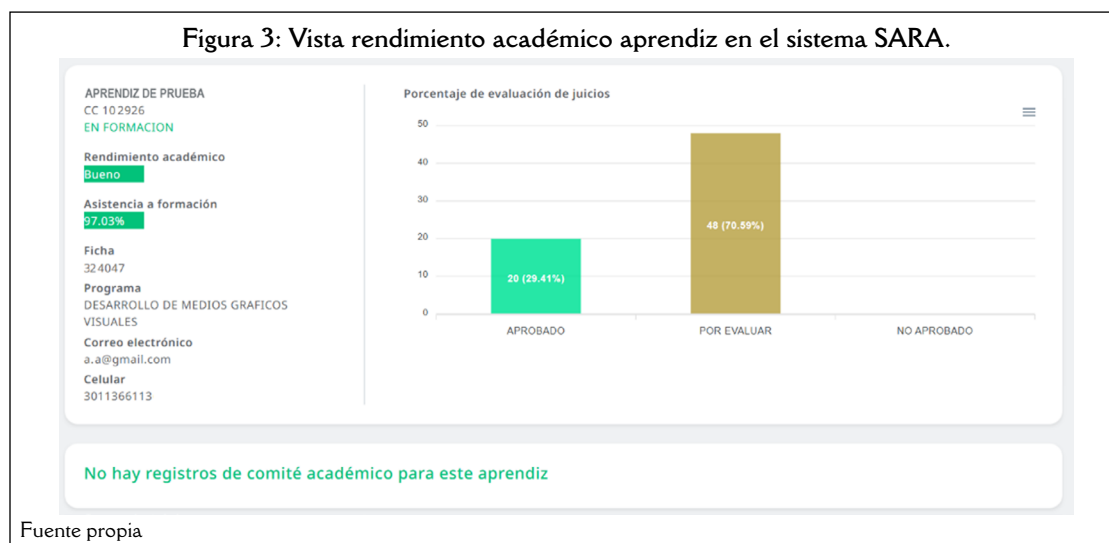
Figura 2: Vista horario instructor en el sistema SARA.

	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
6:00 - 7:00	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Proyecto 2 + Pruebas (Tc.) 2771132 A 803		
7:00 - 8:00	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Proyecto 2 + Pruebas (Tc.) 2771132 A 803		
8:00 - 9:00	CSS, HTML 2821717 A 802	CSS, HTML 2821717 A 802	CSS, HTML 2821717 A 802		CSS, HTML 2821717 A 802	
9:00 - 10:00						
10:00 - 11:00	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Python (Machine Learning + Big Data) 2617472 A 702	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	
11:00 - 12:00	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Python (Machine Learning + Big Data) 2617472 A 702	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	Proyecto 2 + Pruebas (Tc.) 2771132 A 803	

Fuente propia

La base de datos unificada permite tener dos tipos de sistemas unificados los físicos y los químicos. Se importa un archivo adicional que contiene los itinerarios de cada programa de formación con el propósito de cruzar datos y permitir una visión detallada de los resultados de aprendizaje que cada instructor debe orientar en sus respectivos grupos. Para lograr esto, se recorre el horario del instructor día a día, hora por hora, y clase por clase. Para cada clase, se accede al detalle del itinerario correspondiente. Esta información se organiza y almacena en un diccionario, asegurando que cada detalle único sea correctamente agregado. Esto posibilita visualizar tanto el horario como los detalles específicos de cada grupo, como se muestra en la imagen a continuación. (Figura 2)

Por otro lado, se tiene varias funciones que permiten la visualización del rendimiento académico de un aprendiz y su historial en los comités de seguimiento y evaluación. Además, se cuenta con una función que lee datos de un archivo JSON, los procesa y calcula el porcentaje de aprobación para cada estudiante en específico. (Figura 3)



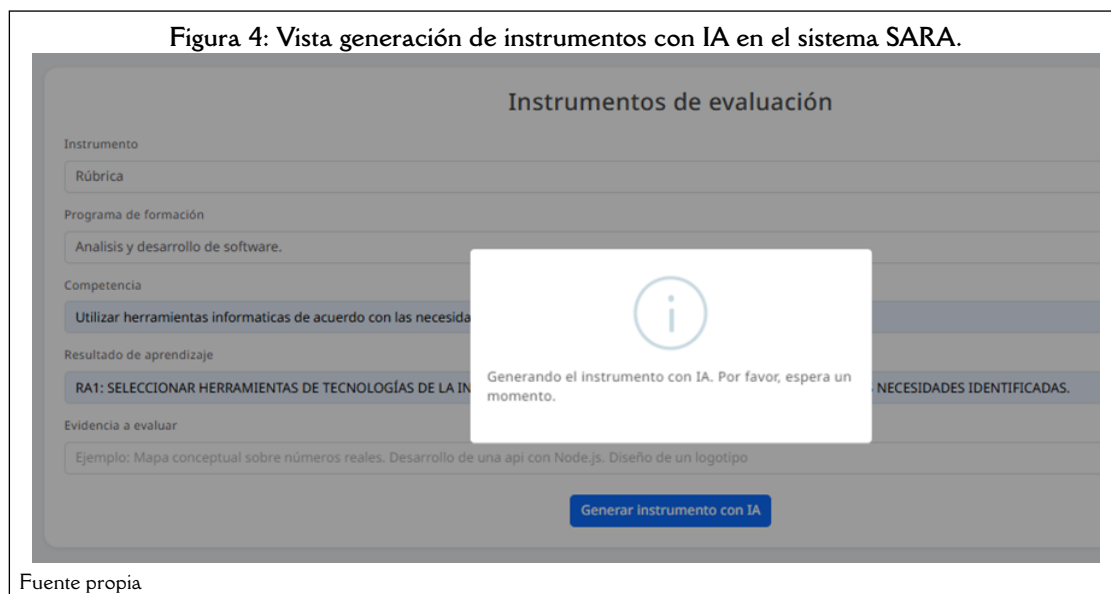
El sistema de información antes descrito, puede ser clasificado como un sistema de información formal ya que su operación está basada en un conjunto de estándares, normas y procedimientos, los cuales permiten la generación de información y que esta llegue a quien lo necesite en el momento oportuno para dinamizar y contribuir con el proceso de programación de aprendices, instructores y ambientes de formación. También puede ser clasificado como un sistema automatizado de información operativo, ya que obedece a un proceso rutinario, que prepara y mantiene los registros de datos originados por la operación elemental (rutinaria) de la organización. A su vez se complementa del tipo de Sistema Gerencial, ya que permite que desde las coordinaciones se puedan tomar decisiones y solución de situaciones propias del proceso de ejecución de la formación profesional integral (FPI).

1.1.1. Inteligencia artificial en el sistema SARA

La inteligencia artificial (IA) se ha consolidado como un eje transformador en la evolución de los sistemas de información, permitiendo el diseño de soluciones más ágiles, inteligentes, adaptativas y eficientes en diversos sectores, incluido el educativo. En términos generales, la IA se refiere al desarrollo de sistemas capaces de realizar tareas que tradicionalmente requieren inteligencia humana, como el aprendizaje, la resolución de problemas, la interpretación del lenguaje natural y la toma de decisiones (Morandín-Ahuerma, 2022).

En el desarrollo del Sistema Automatizado de Reportes Académicos (SARA), la inteligencia artificial desempeña un papel esencial en la generación automatizada de actividades interactivas, entre ellas, cuestionarios y juegos de memoria, adicional, genera de manera automática presentaciones e instrumentos de evaluación, tales como listas de chequeo y rúbricas. A través del procesamiento de lenguaje natural (PLN) y algoritmos de aprendizaje automático, el sistema es capaz de analizar los resultados de aprendizaje

definidos en el currículo y generar reportes de evaluación coherentes con las competencias y desempeños requeridos. Este enfoque no solo reduce significativamente el tiempo invertido por los instructores en el diseño de instrumentos y otras actividades, sino que asegura la estandarización y pertinencia de los mismos frente a los objetivos formativos. A continuación, se visualiza el modulo y mensaje al generar instrumentos de evaluación con IA. (Figura 4)



La automatización del diseño evaluativo mediante IA también posibilita una mayor personalización, adaptando las rúbricas o listas de chequeo a las características del programa, nivel de formación o modalidad de enseñanza. Asimismo, garantiza trazabilidad en las decisiones pedagógicas, ya que cada componente generado puede ser validado y ajustado con base en datos históricos de evaluación y desempeño del aprendiz. Estas innovaciones tecnológicas, integradas en SARA, confirman que la inteligencia artificial puede actuar como un catalizador en la profesionalización de la evaluación educativa, mejorando la equidad, objetividad y eficiencia del proceso. Su implementación representa un avance significativo hacia entornos educativos más inteligentes, con capacidad de respuesta en tiempo real y orientados a la mejora continua.

1.2. Gestión educativa sistematizada

La gestión educativa es un campo especializado que se encarga de organizar, dirigir y optimizar los recursos y procesos dentro de las instituciones educativas para lograr una educación de calidad. Su objetivo principal es crear y mantener entornos que favorezcan la enseñanza y el aprendizaje efectivos, involucrando a todos los actores de la comunidad educativa. Por ende, es un proceso integral y sistemático que comprende la planificación, organización, dirección y control de los recursos humanos, financieros y materiales pedagógicos en las instituciones educativas. Su finalidad es garantizar el logro de los objetivos institucionales y promover una educación de calidad, equitativa y eficiente (Jiménez Riofrío, 2022). En este contexto, la incorporación de tecnologías digitales ha transformado significativamente la manera en que se gestionan dichos procesos, especialmente mediante sistemas automatizados que permiten optimizar la asignación y el uso de recursos en los diferentes niveles organizativos, ya que contribuyen a una planificación más precisa y a una distribución equitativa de la carga laboral entre el personal docente y administrativo.

Asimismo, facilitan la identificación de oportunidades de formación continua, lo que fortalece el desarrollo profesional del talento humano. En cuanto a los recursos materiales y financieros, estas herramientas tecnológicas posibilitan una administración más eficiente, al ofrecer datos precisos y en tiempo real sobre el uso de aulas, equipos y materiales didácticos. Esto permite una toma de decisiones presupuestaria más

informada y una gestión logística más efectiva. Desde una perspectiva estratégica, la gestión educativa automatizada se configura como una herramienta clave para lograr resultados superiores en contextos con limitaciones de tiempo y recursos, elevando la calidad de los servicios ofrecidos a la comunidad académica (Sánchez Armas y Delgado Bardales, 2020).

Además, el auge de técnicas avanzadas como el aprendizaje automático y la inteligencia artificial ha ampliado el alcance de los sistemas de información académica. Estas tecnologías no solo permiten detectar patrones ocultos y predecir comportamientos, sino que también apoyan intervenciones proactivas, como la identificación temprana de estudiantes en riesgo de deserción o bajo rendimiento, facilitando la implementación de estrategias pedagógicas personalizadas.

La integración de las Tecnologías de la Información y la Comunicación (TIC) ha revolucionado la gestión educativa al digitalizar registros, automatizar procesos rutinarios y facilitar el análisis de grandes volúmenes de datos. Este cambio no solo reduce los errores y los tiempos asociados a procedimientos manuales, sino que también libera recursos para ser destinados a actividades de mayor valor estratégico (Briceño Toledo et al., 2020). Es por ello que, explorar el impacto de la automatización en la gestión educativa es fundamental para comprender cómo estas soluciones tecnológicas permiten una administración más eficaz de los recursos, fortaleciendo así la capacidad institucional de responder a los desafíos del sistema educativo en la era digital (Graue et al., 2019).

Es así, como la gestión educativa contemporánea requiere no solo de enfoques administrativos eficientes, sino también de soluciones tecnológicas que apoyen la toma de decisiones estratégicas mediante el análisis inteligente de datos. La implementación de sistemas automatizados como SARA, responde a esta necesidad al integrar tecnologías emergentes como el análisis de datos masivos, la inteligencia artificial y el aprendizaje automático en los procesos de seguimiento académico, planificación docente y administración institucional. De acuerdo con (Aghaei et al., 2023), los sistemas de gestión educativa basados en datos permiten identificar con mayor precisión las necesidades de los aprendices y optimizar los procesos institucionales, lo que conlleva mejoras en la retención estudiantil y la calidad del aprendizaje. En el mismo sentido, la (OECD, 2023) destaca que la digitalización de la educación superior no debe limitarse a la virtualización de clases, sino que debe transformar profundamente la gestión institucional, promoviendo una cultura de innovación y transparencia.

En este contexto, herramientas como SARA representan una oportunidad estratégica para fortalecer la gobernanza académica y aumentar la capacidad adaptativa de las instituciones educativas frente a los desafíos del siglo XXI. Como respuesta concreta a esta transformación digital en la gestión educativa, el sistema SARA ha sido concebido e implementado inicialmente en el SENA, desde la Coordinación de TIC del Centro de Servicios y Gestión Empresarial. Se espera que, por su alcance e impacto descrito en el presente escrito, esta solución pueda ser replicada y adaptada por otras coordinaciones, así como por diferentes Centros de Formación del SENA o instituciones de educación superior en el territorio nacional.

Desde una perspectiva de política institucional, la literatura señala que la optimización de la gestión educativa no depende únicamente de soluciones tecnológicas, sino también de la adopción de lineamientos organizativos y normativos que favorezcan su sostenibilidad. Entre las recomendaciones más citadas están: la implementación de estándares de interoperabilidad para asegurar el intercambio de datos entre sistemas; el establecimiento de políticas de gobernanza de datos, orientadas a calidad, privacidad y trazabilidad; y la formación docente en competencias digitales, como condición necesaria para la apropiación de estas herramientas (European Commission, 2020; Santos et al., 2021). Así, la optimización de la gestión educativa debe entenderse como un proceso integral en el que convergen tecnología, organización y política institucional. Este estudio contribuye a esa perspectiva al presentar y evaluar SARA, un sistema automatizado que, además de evidenciar factibilidad técnica y aceptación por parte de los instructores, formula recomendaciones de política para garantizar su adopción y replicabilidad.

En coherencia con lo anterior, dentro de los resultados se realizó un análisis de bibliometría que permitió identificar y examinar estudios publicados en revistas indexadas en Scopus, los cuales, representan aportes recientes y relevantes para la temática. La revisión de la literatura se sintetiza en la siguiente tabla, donde se observa cómo dichos estudios, tanto de carácter teórico como empírico, aportan evidencia para fundamentar la pertinencia del sistema automatizado de reportes académicos en la gestión educativa del SENA. (Tabla 1)

Tabla 1: Comparación de literatura con la pertinencia del sistema SARA

Tema / Eje de análisis	Principales hallazgos	Pertinencia para el sistema automatizado de reportes académicos	Fuentes / Autores
Optimización en la Gestión de información	Los estudios muestran cómo la recopilación sistemática y estandarizada de datos permite optimizar la gestión de la información, en estos casos teóricos, del área de salud y apoyar la toma de decisiones.	El enfoque en datos masivos y análisis longitudinales sustenta la necesidad de sistemas de información automatizados también en la gestión de educación.	(Brauer et al., 2024; Feigin et al., 2024; Ferrari et al., 2024; Mensah et al., 2023; Naghavi et al., 2024; Ng et al., 2025; Schumacher et al., 2024; Steinmetz et al., 2024; Van Gelder et al., 2024; Wu et al., 2025)
Inteligencia artificial y aprendizaje automático	Los autores identifican que la IA y el aprendizaje profundo optimizan análisis de datos, predicción de patrones y generación de reportes automáticos.	Apoya la inclusión de algoritmos de IA en los sistemas académicos para mejorar precisión, personalización y análisis predictivo.	(Chen et al., 2024; Ma et al., 2021; Mienye y Sun, 2022, 2023; Sarker, 2022; Sarker et al., 2021; Su et al., 2022; Zhou et al., 2024; Zhu et al., 2023)
Tecnologías digitales, industria 4.0, Ciberseguridad y metaverso	Se resalta la importancia de la digitalización, automatización y tecnologías inmersivas para mejorar procesos y experiencias. Adicional que los reportes automatizados deben cumplir con estándares de privacidad y seguridad de datos sensibles.	Refuerza la idea de incorporar entornos digitales avanzados en la gestión académica y administrativa. En la ciberseguridad, se relaciona directamente con la gestión académica al manejar información personal de los aprendices.	(Javaid et al., 2024; Javaid et al., 2022, 2023; Liu et al., 2022; Park y Kim, 2022)
Educación, comunicación y sociedad digital	Los autores enfatizan que la participación de usuarios, la accesibilidad y la confianza en las tecnologías son claves para la adopción de sistemas automatizados.	Fundamenta que la adopción de un sistema de reportes debe considerar usabilidad, motivación y pertinencia para aprendices e instructores y gobernanza institucional.	(Haleem et al., 2022; Wu et al., 2023; Wu et al., 2024; Yu et al., 2022)

Fuente propia

los estudios revisados confirman que la integración de tecnologías digitales, modelos de análisis de datos y lineamientos organizativos constituye un eje fundamental para avanzar en la optimización de la gestión educativa. La evidencia recopilada muestra que los hallazgos internacionales no solo resaltan la viabilidad técnica de sistemas automatizados, sino también la necesidad de acompañarlos con políticas institucionales, formación docente y marcos de gobernanza de datos. De esta manera, la propuesta del sistema SARA se enmarca en un escenario de validez académica y pertinencia práctica, al articular tendencias globales con las necesidades específicas de la educación en instituciones de formación profesional.

1.3. Análisis de sentimiento

El análisis de sentimiento son procesos fundamentales para la evaluación cualitativa, especialmente en escenarios donde es crucial comprender la percepción de los usuarios, para este caso en relación con el desempeño, la usabilidad y el valor añadido del sistema SARA. Esta técnica permite transformar datos no estructurados, como opiniones, entrevistas y respuestas abiertas, en información significativa que puede guiar procesos de mejora continua en entornos organizacionales y educativos. El análisis de sentimiento es una técnica derivada del procesamiento del lenguaje natural (PLN), orientada a identificar, extraer y clasificar emociones y valoraciones expresadas en el lenguaje escrito y corporal, atribuyéndoles una polaridad (positiva, negativa o neutra) (Quintero López et al., 2024). En el contexto de validación de sistemas de información, esta técnica permite evaluar de manera sistemática el nivel de satisfacción del usuario, detectar puntos críticos en la experiencia de uso y fundamentar decisiones estratégicas para la optimización del sistema. En el análisis de lenguaje escrito se abarca un conjunto más amplio de técnicas de minería de datos y PLN que permiten identificar patrones temáticos, estructuras semánticas y tendencias lingüísticas en grandes volúmenes de texto. Esta aproximación resulta particularmente útil para analizar datos cualitativos derivados de entrevistas, foros institucionales o registros abiertos, proporcionando una comprensión más profunda del discurso de los usuarios (Campos-Malpartida et al., 2024).

Para este análisis se empleó el modelo `nlptown/bert-base-multilingual-uncased-sentiment`, una arquitectura de aprendizaje profundo basada en BERT (Bidirectional Encoder Representations from Transformers), entrenada para tareas de clasificación de sentimientos en múltiples idiomas, incluido el español. Este modelo ha sido optimizado para analizar texto corto, como respuestas abiertas, comentarios y frases expresadas en lenguaje natural, asignando una puntuación de sentimiento en una escala ordinal que va de 1 (muy negativo) a 5 (muy positivo). La elección de este modelo responde a varias ventajas fundamentales. En primer lugar, su naturaleza multilingüe lo hace especialmente útil en contextos latinoamericanos, donde el idioma predominante es el español, pero pueden coexistir términos técnicos y expresiones locales diversas. En segundo lugar, su capacidad de trabajar con textos sin distinción de mayúsculas o minúsculas (`uncased`) mejora la robustez del análisis frente a inconsistencias comunes en la redacción de los usuarios.

El modelo fue implementado utilizando la biblioteca `Transformers` de Hugging Face, lo que permitió integrarlo de manera eficiente en el flujo de procesamiento del sistema automatizado. Una vez transcritas y limpiadas las entrevistas semiestructuradas, cada fragmento textual relevante fue sometido al modelo, que devolvió una puntuación de sentimiento para cada unidad de análisis. Estas puntuaciones fueron luego interpretadas en función de su polaridad (positiva, neutra o negativa), lo que permitió identificar patrones emocionales en las percepciones de los instructores respecto al Sistema Automatizado de Reportes Académicos (SARA). Además, este modelo se beneficia del preentrenamiento con grandes volúmenes de datos multilingües, lo que mejora significativamente la precisión del análisis en dominios no especializados, como el educativo. Su uso refuerza el enfoque metodológico del estudio al complementar el análisis cualitativo tradicional con una herramienta de inteligencia artificial capaz de sistematizar y escalar la interpretación emocional del discurso docente.

2. Metodología

Este estudio parte de la identificación de vacíos en la gestión académica del SENA, relacionados con la coexistencia de múltiples fuentes de datos, el predominio de procesos manuales y la ausencia de mecanismos de interoperabilidad, lo que limita la trazabilidad de la información y retrasa la toma de decisiones. En este marco, la investigación aporta un modelo innovador llamado SARA, que responde a esta problemática. La metodología corresponde a una investigación aplicada, con enfoque mixto y alcance descriptivo-exploratorio, se estructuró en dos componentes complementarios: en primer lugar, el desarrollo técnico del Sistema Automatizado de Reportes Académicos (SARA), que combina una descripción de la arquitectura tecnológica escalable y un análisis bibliométrico de la literatura existente; y en segundo lugar, un análisis de percepciones de instructores, apoyado en procesamiento de lenguaje natural. Esta integración constituye una contribución novedosa al abordar la gestión educativa desde una perspectiva interdisciplinaria, articulando soluciones técnicas y pedagógicas.

2.1. Desarrollo técnico y análisis bibliométrico

El desarrollo técnico de SARA se basó en la arquitectura Modelo-Vista-Controlador (MVC), seleccionada por su capacidad de garantizar escalabilidad, mantenibilidad y separación de responsabilidades entre componentes. El sistema fue programado en Python, aprovechando bibliotecas especializadas como `Pandas` para el procesamiento de datos, y `NLTK` y `Scikit-learn` para análisis de lenguaje natural y de sentimientos.

Para el almacenamiento, se implementó una base de datos `NoSQL`, debido a la naturaleza heterogénea de los registros académicos (Excel, reportes de `SOFIA Plus` y archivos `JSON`), que superan los 140.000 registros y presentan estructuras semiestructuradas. Esta solución permitió realizar procesos de extracción, transformación y limpieza de datos (ETL), reducir duplicidades y garantizar la integración eficiente de información. El sistema incorpora módulos de visualización que permiten organizar horarios de instructores y aprendices, consultar itinerarios de programas de formación y analizar el rendimiento académico de los estudiantes. De esta manera, SARA opera como un sistema de información automatizado, orientado tanto al soporte operativo como a la toma de decisiones gerenciales en el marco de la formación profesional integral.

El análisis bibliométrico es un método cuantitativo para estudiar la literatura científica utilizando datos numéricos y estadísticos para identificar patrones en publicaciones, citas, autores y tendencias de investigación. Con el fin de contextualizar la relevancia de este desarrollo en la gestión educativa, se realizó un análisis en la base de datos `Scopus` elegida por criterios de cobertura, calidad y pertinencia académica. Para la búsqueda se diseñaron tres ecuaciones: `artificial intelligence AND cluster`; `educational management OR`

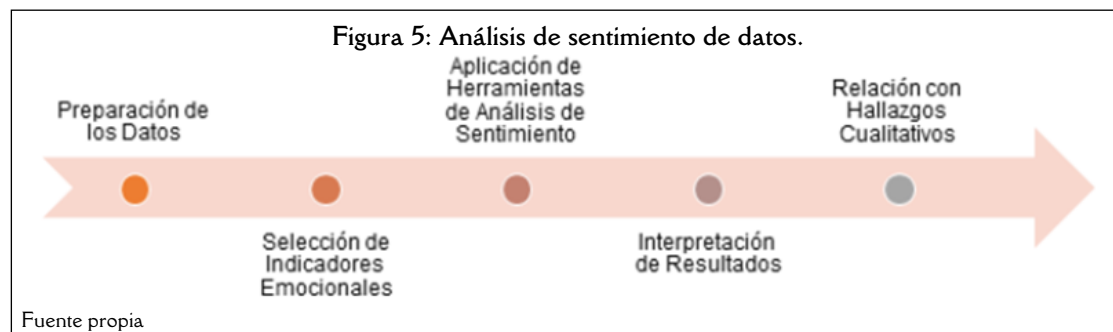
learning analytics; y pubyear > 2020 AND pubyear < 2027. En el filtro se utilizó rango de años entre 2021 al 2026. La consulta arrojó una muestra total de 4.922 documentos, que posteriormente fueron procesados en el software VOSviewer®, aplicando análisis de coautoría. Para el mapeo se aplicó los filtros de type of analysis: co-authorship; counting method: full counting; unit of analysis: authors, y se seleccionaron los documentos por número de citas, se descartaron los que no han sido citados y resultados repetidos por las diferentes búsquedas. Este procedimiento permitió identificar las principales redes de colaboración científica, autores con mayor producción y tendencias emergentes en torno al uso de inteligencia artificial, gestión automatizada y analítica de aprendizaje, dando como resultado 28 documentos. Dichos hallazgos sirvieron para contextualizar el problema de investigación, evidenciar vacíos en la literatura reciente y fundamentar la pertinencia del desarrollo e implementación de SARA.

2.2. percepción del sistema SARA

El análisis cualitativo estuvo orientado a explorar y comprender en profundidad las percepciones, valoraciones y experiencias de instructores en relación con la implementación del Sistema Automatizado de Reportes Académicos (SARA) como herramienta para optimizar la gestión educativa. Este enfoque resulta pertinente para captar la complejidad del fenómeno en su contexto natural, así como para valorar aspectos subjetivos como la usabilidad, la utilidad percibida y las emociones asociadas al uso del sistema. La muestra está compuesta por 7 instructores, cuya experiencia y conocimiento en el campo educativo es representativo y significativo dentro cuerpo académico del SENA, Centro de Servicios y Gestión empresarial. La muestra es seleccionada de manera no probabilística, considerando la diversidad de perspectivas y experiencias. Se escogieron aquellos que cumplen algunos de los siguientes criterios: Instructor Sena – líder de área – Antigüedad mayor a 5 años de experiencia en el SENA – Instructor con menos de un año en el SENA - instructores con experiencia directa en el uso del sistema SARA.

Para la recolección y análisis de datos se utilizó como instrumento la entrevista semiestructurada. Esta técnica permite una exploración profunda de las percepciones y experiencias de los participantes, así como la flexibilidad necesaria para seguir líneas de indagación emergentes durante la entrevista. El instrumento se diseñó considerando preguntas abiertas que permiten a los participantes expresar libremente sus opiniones y experiencias en relación con la implementación del sistema automatizado de reportes académicos. Se abordaron temas como la eficacia del sistema, los desafíos enfrentados durante su implementación, las ventajas y desventajas percibidas, y las recomendaciones para su mejora. Las entrevistas se hicieron de manera individual y en un entorno propicio para la reflexión y el diálogo, procurando establecer una relación de confianza con los participantes para fomentar la apertura y la sinceridad en sus respuestas.

El análisis de los datos recopilados se llevó a cabo en el modelo nlptown/bert-base-multilingual-uncased-sentiment considerando el aspecto cualitativo de las respuestas de los participantes en las entrevistas. Se utilizarán técnicas de análisis de sentimiento para identificar las emociones y actitudes expresadas por los instructores en relación a la implementación del sistema automatizado SARA. En la siguiente Figura 5 se resumen los pasos que se llevaron a cabo para el análisis de sentimiento.



Antes de realizar el análisis de sentimiento, se lleva a cabo la transcripción y codificación de las entrevistas. Se inicia con la preparación de los datos, eliminando los stopwords que son palabras que se omiten en las respuestas por que no aportan un sentimiento o una valoración que signifique positivo o negativo. Solo

se identifica unidades de texto relevantes para el análisis, como opiniones, comentarios y experiencias relacionadas con el sistema. Luego se define indicadores emocionales relevantes para el estudio, como palabras clave asociadas con emociones positivas, por ejemplo, “eficaz”, “útil” y emociones negativas, por ejemplo, “frustración”, “dificultad”. Estos indicadores guiarán el proceso de análisis de sentimiento. En la aplicación de herramientas de análisis de sentimiento se utiliza herramientas como algoritmos de aprendizaje automático o técnicas de procesamiento de lenguaje natural, para evaluar el tono emocional de los datos. Estas herramientas asignan un valor de sentimiento a cada unidad de texto identificada en las entrevistas. Los resultados del análisis de sentimiento se interpretan en función de la polaridad emocional identificada en las respuestas de los participantes. Se analizan las emociones predominantes expresadas hacia el sistema automatizado de reportes académicos, así como las variaciones individuales entre los docentes expertos. Finalmente, se establece una relación entre los resultados del análisis de sentimiento y los hallazgos cualitativos obtenidos durante el análisis de datos. Se explorarán posibles correlaciones entre las emociones expresadas y los temas emergentes identificados en las entrevistas, proporcionando una comprensión más completa de las percepciones de los participantes.

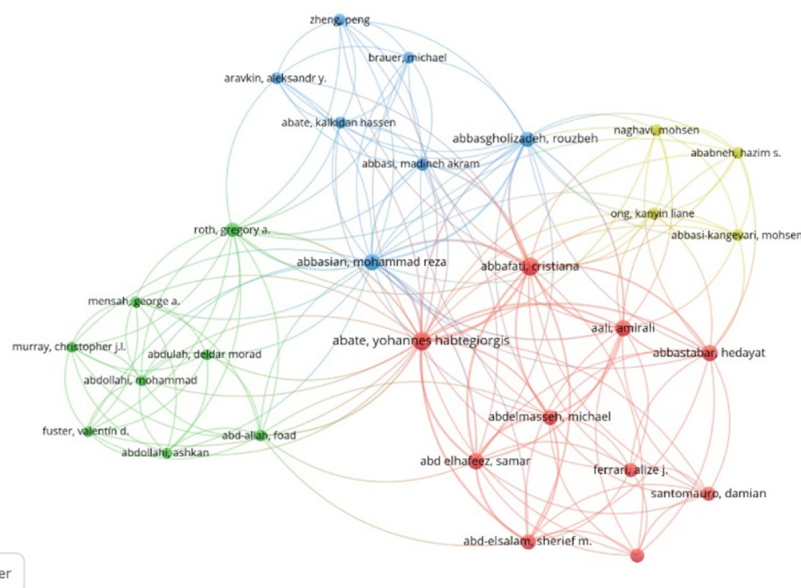
3. Resultados

3.1. Análisis de coautoría

Con el fin de identificar las principales redes de colaboración científica en el campo, se realizó un análisis de coautoría a partir de la base de datos Scopus, procesada con el software VOSviewer® (Fig. 6). Este análisis permite visualizar cómo se estructuran las relaciones entre autores a partir de publicaciones conjuntas y refleja el grado de interconexión en la producción académica reciente. En la red obtenida, cada nodo representa a un autor, cuyo tamaño indica el número de publicaciones y su relevancia dentro del conjunto analizado. Las líneas muestran los vínculos de coautoría, es decir, la frecuencia con la que los investigadores han publicado en colaboración. A su vez, los nodos se agrupan en clústeres diferenciados por colores, que corresponden a comunidades de autores que colaboran de manera más recurrente entre sí.

La Figura 6 evidencia al menos tres clústeres principales que agrupan a los autores más productivos y con mayor conexión en el ámbito de la inteligencia artificial aplicada a la educación, la gestión de datos y la analítica de aprendizaje. Esta distribución sugiere que el campo se estructura en redes interconectadas, pero con núcleos de especialización que marcan las tendencias investigativas actuales.

Figura 6: Análisis de coautoría.



Fuente propia

La visualización muestra al menos tres clústeres principales diferenciados por colores, lo que refleja la existencia de redes consolidadas de investigación. El clúster rojo concentra a los autores con mayor densidad de conexiones, lo que sugiere su papel central en la producción y diseminación de conocimiento en este campo. El clúster verde agrupa a investigadores que, aunque menos interconectados, presentan aportes relevantes en subtemáticas específicas. Por su parte, el clúster azul y los nodos periféricos evidencian colaboraciones emergentes, indicando líneas de investigación que podrían fortalecerse en el futuro. En conjunto, esta red demuestra que la literatura sobre el tema no se desarrolla de manera aislada, sino en torno a comunidades científicas interdependientes, lo que otorga solidez al marco de referencia utilizado en este estudio. La distribución de clústeres evidencia núcleos de especialización y cierta fragmentación temática. Estos patrones se corresponden con la literatura reciente, que subraya la importancia de articular innovaciones tecnológicas con prácticas pedagógicas bajo marcos de interoperabilidad, gobernanza de datos y sostenibilidad digital. El análisis bibliométrico evidenció que estas temáticas se posicionan como líneas de investigación emergentes y prioritarias en el ámbito de la gestión educativa, especialmente en estudios que abordan la analítica de aprendizaje y la digitalización institucional. La fragmentación observada entre clústeres de autores refuerza la necesidad de propuestas integradoras que reduzcan las brechas entre lo técnico y lo pedagógico.

Para garantizar la adopción del sistema SARA, es recomendable la interoperabilidad de sistemas, avanzando en la adopción de estándares que permitan la integración de plataformas como SOFIA Plus, reduciendo la fragmentación de información y optimizando los flujos de datos académicos. Por otro lado, establecer lineamientos claros sobre calidad, privacidad y trazabilidad de la información académica, garantizando que los registros consolidados en SARA cumplan con principios de seguridad y confiabilidad. Diseñar programas permanentes de capacitación para instructores y coordinadores, orientados no solo al uso de SARA, sino también al fortalecimiento de habilidades en analítica de datos y gestión digital. Finalmente, proyectar la implementación progresiva de SARA en otros centros de formación del SENA e instituciones de formación profesional, evaluando su impacto en la reducción de cargas administrativas y en la toma de decisiones académicas.

En consecuencia, el desarrollo de SARA se plantea como una contribución práctica que responde a estas demandas, al centralizar información académica dispersa, facilitar la trazabilidad de los aprendizajes y optimizar la toma de decisiones en entornos de formación profesional. Su diseño escalable y replicable lo convierte en una herramienta con potencial de ser adoptada en otros contextos educativos, aportando evidencia aplicada sobre cómo la convergencia entre tecnología, pedagogía y política institucional puede fortalecer la gestión académica.

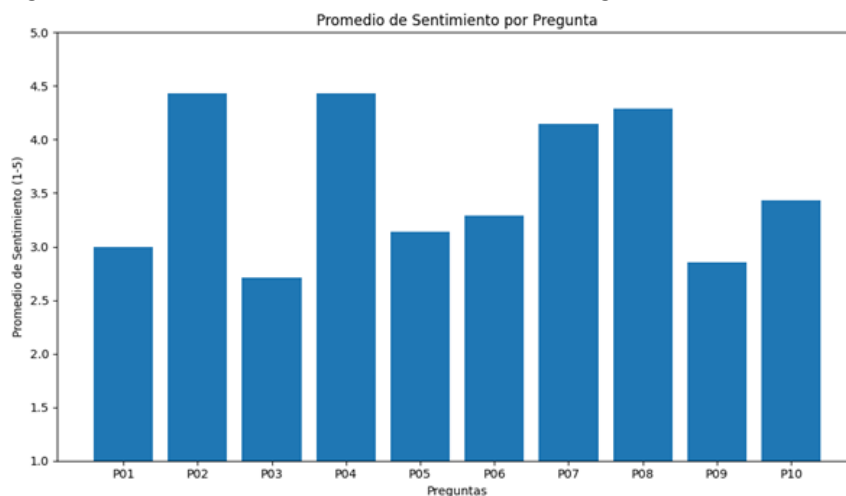
3.2 Análisis de sentimientos

Se aplicó un análisis de sentimientos a las respuestas obtenidas en entrevistas con siete instructores, a partir de un cuestionario compuesto por diez preguntas orientadas a explorar su experiencia antes y después de la implementación de SARA. El instrumento permitió recoger percepciones, identificar desafíos y recopilar recomendaciones sobre el uso del sistema. Las preguntas fueron diseñadas para evaluar de manera específica las valoraciones y experiencias de los participantes respecto al Sistema Automatizado de Reportes Académicos (SARA).

- P01. Experiencia antes del sistema automatizado de reportes académicos SARA.
- P02. ¿Podría describir brevemente su experiencia con la herramienta SARA?
- P03. ¿Qué desafíos ha experimentado al utilizar esta herramienta?
- P04. ¿Describa su percepción frente al uso de la herramienta en el proceso de programación de instructor y seguimiento de los aprendices?
- P05. Frente al proceso de programación de instructor, ¿qué le gustaría que tuviera el sistema SARA?
- P06. Frente al proceso de seguimiento de aprendices, ¿qué le gustaría que tuviera el sistema SARA?
- P07. Comente su percepción frente a la implementación de SARA en el proceso y los tiempos en su rol de instructor desarrollo.
- P08. ¿Cuáles son las ventajas percibidas en el uso de la herramienta?
- P09. ¿Cuáles son las desventajas percibidas en el uso de la herramienta?
- P10. Recomendaría el uso de la herramienta para implementarla en otras coordinaciones y procesos. ¿Por qué?

Para este análisis se empleó un modelo multilingüe de aprendizaje automático (nlptown/bert-base-multilingual-uncased-sentiment) capaz de identificar sentimientos en textos en español, el cual clasifica las respuestas en una escala de 1 a 5, donde 1 representa un sentimiento muy negativo y 5 un sentimiento muy positivo. (Figura 7)

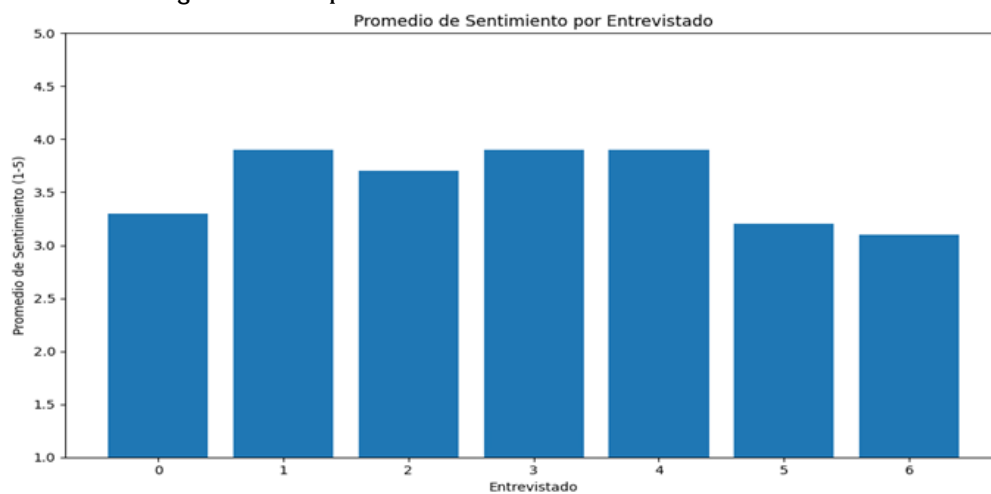
Figura 7: Valor promedio del sentimiento de cada pregunta de las entrevistas.



Fuente propia

A partir del análisis de los resultados promedio de los sentimientos asociados a cada una de las preguntas, se observa una tendencia general positiva, con especial énfasis en las preguntas P02, P04, P07 y P08, todas ellas con promedios superiores a 4, lo que evidencia una valoración muy favorable por parte de los instructores. No obstante, se identifican áreas de mejora en las preguntas P03 y P09, cuyos promedios se aproximan o se sitúan por debajo de 3, sugiriendo percepciones más críticas en aspectos relacionados con los desafíos y desventajas del sistema. En conjunto, el hecho de que la mayoría de los promedios estén por encima de 3 indica una estabilidad emocional en la respuesta y una experiencia globalmente positiva con la herramienta SARA. La siguiente figura presenta los resultados promedio de los puntajes de sentimientos por entrevistado que tras responder cada pregunta. En total son 7 entrevistados los cuales son etiquetados de 0 a 6. (Figura 8)

Figura 8: Valor promedio del sentimiento de cada entrevistado.



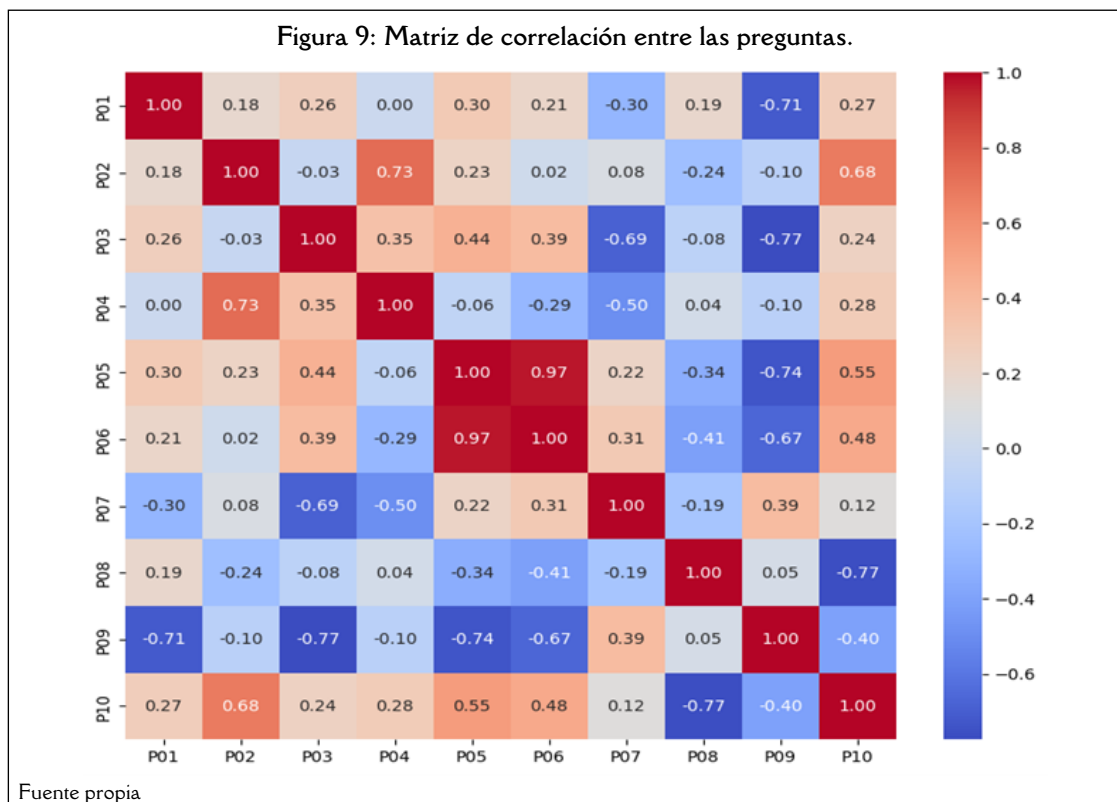
Fuente propia

En general, se observa una tendencia positiva y estable, con la mayoría de los entrevistados (1, 2, 3 y 4) mostrando promedios de sentimiento cercanos a 4, lo que sugiere una valoración favorable y homogénea del sistema SARA en esos casos. El entrevistado 0 también mantiene una percepción levemente positiva con un promedio de 3.3, mientras que los entrevistados 5 y 6 muestran una actitud más crítica, con promedios de 3.2 y 3.1 respectivamente, aunque aún dentro del rango medio-positivo. Esta distribución refleja una experiencia general satisfactoria, con oportunidades específicas de indagar más en los casos donde las valoraciones fueron más bajas para entender posibles dificultades o expectativas no cumplidas.

A continuación, se presenta ahora una matriz de correlación para ver qué tan relacionados están los sentimientos registrados por cada pregunta en las entrevistas. Los valores van de -1 a 1:

- 1 significa una relación perfecta y directa: cuando alguien responde alto en una pregunta, también lo hace en la otra.
- -1 significa una relación perfecta pero inversa: si responde alto en una, responde bajo en la otra.
- 0 significa que no hay relación clara entre las dos preguntas.

A partir de eso, se pueden observar las siguientes relaciones importantes: (Figura 9)



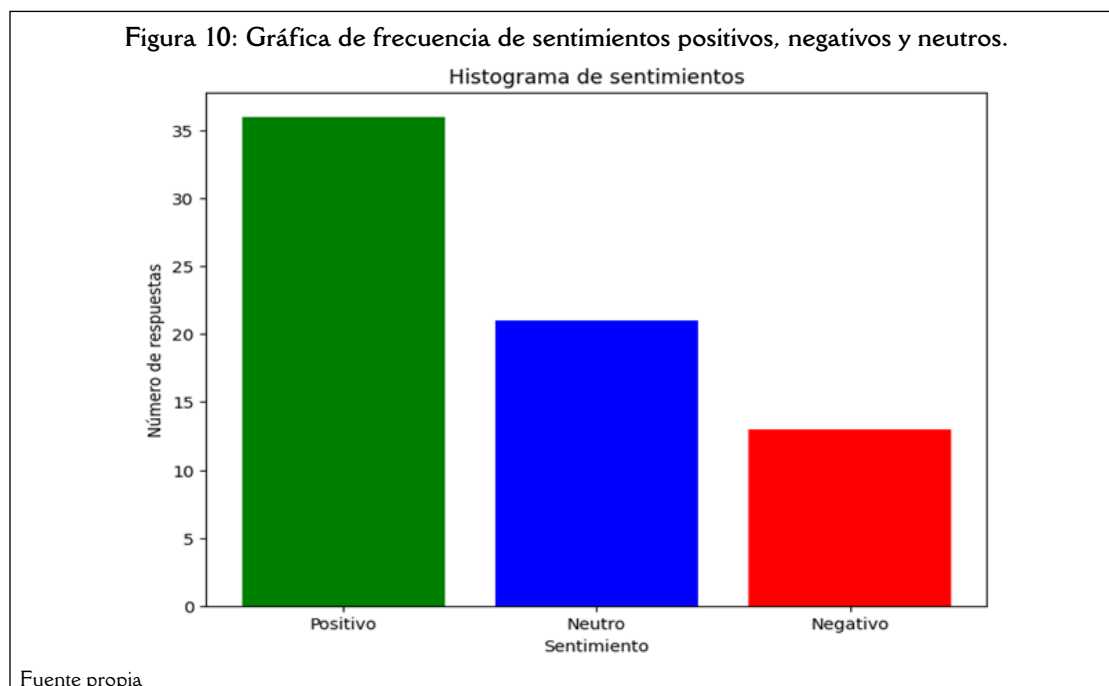
Se observa una relación casi perfecta entre las preguntas P05 y P06 (0.97), lo que indica que quienes tienen una opinión positiva sobre cómo debería mejorar el proceso de programación de instructores también valoran positivamente el seguimiento de aprendices. Asimismo, hay una fuerte relación entre P02 y P04 (0.73), y entre P02 y P10 (0.68), lo que sugiere que quienes reportan una buena experiencia general con la herramienta tienden a percibirla como útil en el proceso y están más dispuestos a recomendarla. También destaca la relación entre P05 y P10 (0.55), donde las personas que proponen mejoras específicas para el sistema suelen estar más dispuestas a recomendar su implementación.

Se identifican relaciones negativas destacadas entre varias preguntas, lo que sugiere percepciones críticas o contrastantes. La más fuerte es entre P03 y P09 (-0.77), donde quienes enfrentan más desafíos con la herramienta también perciben más desventajas. De forma inversa, quienes tuvieron una experiencia

más negativa antes del sistema (P01) tienden a ver menos desventajas en la actualidad (P09), con una correlación de -0.71 . Además, quienes reportan mayores desafíos (P03) tienden a tener una percepción más negativa sobre la implementación y los tiempos de uso de la herramienta (P07), con una correlación de -0.69 . Finalmente, resulta llamativo que quienes perciben más ventajas en el uso de la herramienta (P08) no siempre son quienes la recomiendan (P10), con una correlación de -0.77 , lo que podría deberse a factores externos como políticas institucionales.

3.3. Análisis de texto

Con el propósito de comprender la percepción de los usuarios frente al sistema SARA, se realizó una clasificación de las respuestas recolectadas, segmentándolas en tres categorías de sentimiento: Positivo, Neutro y Negativo. Esta categorización se basó en los valores asignados por los encuestados, considerando como positivas aquellas respuestas con una valoración superior a 3, neutras las que recibieron un valor exacto de 3, y negativas las que obtuvieron una puntuación inferior a dicho umbral. (Figura 10)



A partir de esta clasificación, se procedió al análisis de las respuestas positivas, negativas y neutras, empleando herramientas visuales como nubes de palabras para identificar los términos más frecuentes y dendrogramas para explorar similitudes entre las respuestas.

Un dendrograma es un tipo de diagrama que permite visualizar agrupamientos jerárquicos entre elementos en este caso, respuestas textuales, mostrando cómo se relacionan entre sí con base en su similitud semántica. Esta herramienta es especialmente útil en el análisis cualitativo para identificar clústeres o grupos temáticos emergentes, ya que proporciona una representación visual de las distancias o similitudes entre fragmentos de texto, facilitando la interpretación de patrones discursivos complejos.

3.3.1 Análisis de Respuestas Positivas

A continuación, se presenta un análisis de texto enfocado en las respuestas positivas recogidas durante las entrevistas sobre la herramienta SARA. Este análisis busca identificar las palabras más representativas y los patrones comunes en las percepciones favorables de los participantes. Para ello, se incluye una nube de palabras y un dendrograma de similitud, dos herramientas visuales que permiten explorar tanto la frecuencia como la relación entre términos clave. (Figura 11)

E0 - P01 y E6 - P01— que comparten temas similares y podrían reflejar limitaciones sistemáticas de la plataforma que afectan a múltiples usuarios de manera comparable.

4. Conclusiones

Los hallazgos de este estudio evidencian que SARA constituye una herramienta viable para optimizar la gestión académica en instituciones de formación profesional, al reducir la carga administrativa, mejorar la trazabilidad de los aprendizajes y facilitar la toma de decisiones oportunas. Estos resultados coinciden con investigaciones previas sobre analítica de aprendizaje y automatización educativa, que han demostrado mejoras en la predicción de la deserción y en la eficiencia de los procesos institucionales. Más allá del caso específico del SENA, la implementación de sistemas similares debe considerar criterios como: la adopción de estándares de interoperabilidad para garantizar el intercambio seguro de datos, el fortalecimiento de la gobernanza institucional en torno a calidad, privacidad y trazabilidad de la información, y la capacitación docente en competencias digitales para asegurar la apropiación de estas herramientas.

El impacto esperado se sitúa principalmente en el ámbito educativo, al favorecer una gestión basada en evidencia y una mejora en la experiencia de aprendizaje; y secundariamente en el de los medios y la información, al demostrar cómo arquitecturas de datos y técnicas de procesamiento del lenguaje natural pueden aplicarse a la educación de manera interdisciplinaria.

Finalmente, se recomienda escalar el modelo a otros centros de formación, realizar estudios comparativos con instituciones de educación superior y articular estas iniciativas con políticas nacionales de digitalización educativa. De esta forma, se consolidaría un ecosistema de gestión académica más eficiente, sostenible y alineado con los retos actuales de la educación digital.

Referencias

- Aghaei, S., Shahbazi, Y., Pirbabaei, M. y Beyti, H. (2023). A hybrid SEM-neural network method for modeling the academic satisfaction factors of architecture students. *Computers and Education: Artificial Intelligence*, 4, 100122. <https://doi.org/10.1016/j.caeai.2023.100122>
- Arias Ortiz, E., Eusebio, J., Pérez Alfaro, M., Vásquez, M. y Zoido, P. (2021). *Los Sistemas de Información y Gestión Educativa (SIGED) de América Latina y el Caribe: la ruta hacia la transformación digital de la gestión educativa*. Inter-American Development Bank. <https://doi.org/10.18235/0003345>
- Brauer, M., Roth, G. A., Aravkin, A. Y., Zheng, P., Abate, K. H., Abate, Y. H., Abbafati, C., et al. (2024). Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*, 403(10440), 2162-2203. [https://doi.org/10.1016/S0140-6736\(24\)00933-4](https://doi.org/10.1016/S0140-6736(24)00933-4)
- Briceño Toledo, M., Correa Castillo, S., Valdés Montecinos, M. y Hadweh Briceño, M. (2020). Modelo de gestión educativa para programas en modalidad virtual de aprendizaje. *Revista de Ciencias Sociales*, 26(2), 286-298. <https://doi.org/10.31876/rcs.v26i2.32442>
- Camones Gonzales, F. C., Sihuay Fernandez, M. T., Nolberto Sifuentes, V. A. y Padilla Caballero, J. E. A. (2024). Minería de datos: Un enfoque perspectivo desde el contexto educativo. *Revista Tribunal*, 4(9), 138-160. <https://doi.org/10.59659/revistatribunal.v4i9.70>
- Campos-Malpartida, J., Ordoñez-Ramos, E. y Huaylla-Quispe, A. (2024). Análisis del Sentimiento con NLP en tutoría académica: la vocación como factor de mejora en el rendimiento académico. *C&T Riqchary Revista de Investigación en Ciencia y Tecnología*, 6(2), 7-13. <https://doi.org/10.57166/riqchary.v6.n2.2024.123>
- Chen, C., Zheng, F., Cui, J., Cao, Y., Liu, G., Wu, J. y Zhou, J. (2024). Survey and open problems in privacy-preserving knowledge graph: merging, query, representation, completion, and applications. *International Journal of Machine Learning and Cybernetics*, 15(8), 3513-3532. <https://doi.org/10.1007/s13042-024-02106-6>
- European Commission. (2020). *Digital Education Action Plan 2021–2027*. Publications Office of the European Union. <https://education.ec.europa.eu/focus-topics/digital-education/actions>
- Feigin, V. L., Abate, M. D., Abate, Y. H., Abd ElHafeez, S., Abd-Allah, F., Abdelalim, A., Abdelkader, A., et al. (2024). Global, regional, and national burden of stroke and its risk factors, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet Neurology*, 23(10), 973-1003. [https://doi.org/10.1016/S1474-4422\(24\)00369-7](https://doi.org/10.1016/S1474-4422(24)00369-7)
- Ferrari, A. J., Santomauro, D. F., Aali, A., Abate, Y. H., Abbafati, C., Abbastabar, H., Abd ElHafeez, S., et al. (2024). Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*, 403(10440), 2133-2161. [https://doi.org/10.1016/S0140-6736\(24\)00757-8](https://doi.org/10.1016/S0140-6736(24)00757-8)
- García-Peñalvo, F. J. y Corell, A. (2020). La COVID-19: ¿enzima de la transformación digital de la docencia o reflejo de una crisis metodológica y competencial en la educación superior? *Campus Virtuales*, 9(2), 83-98. <https://hdl.handle.net/10366/144140>
- Graue, E., Martuscelli, J. y Martínez Leyva, C. (2019). Educación superior, el futuro del trabajo y la automatización. *Universidades*, (81), 65-76. <https://doi.org/10.36888/udual.universidades.2019.81.38>
- Haleem, A., Javaid, M., Qadri, M. A. y Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, 3, 275-285. <https://doi.org/10.1016/j.susoc.2022.05.004>

- Holmes, W., Bialik, M. y Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning* (2nd ed.). Center for Curriculum Redesign. <https://www.researchgate.net/publication/332180327>
- Javaid, M., Haleem, A., Singh, R. P. y Sinha, A. K. (2024). Digital economy to improve the culture of industry 4.0: A study on features, implementation and challenges. *Green Technologies and Sustainability*, 2(2), 100083. <https://doi.org/10.1016/j.grets.2024.100083>
- Javaid, M., Haleem, A., Singh, R. P. y Suman, R. (2022). Enabling flexible manufacturing system (FMS) through the applications of industry 4.0 technologies. *Internet of Things and Cyber-Physical Systems*, 2, 49-62. <https://doi.org/10.1016/j.iotcps.2022.05.005>
- Javaid, M., Haleem, A., Singh, R. P. y Suman, R. (2023). Towards insighting cybersecurity for healthcare domains: A comprehensive review of recent practices and trends. *Cyber Security and Applications*, 1, 100016. <https://doi.org/10.1016/j.csa.2023.100016>
- Jiménez Riofrio, S. M. (2022). Gestión educativa y liderazgo educativo; las tic en la mejora de la competitividad. *Neutrosophic Computing and Machine Learning*, 19, 66-74. <https://fs.unm.edu/NCML2/index.php/112/article/view/184>
- Kung, C. A. L. L., Noriega, E. R., Huarmiyuri, A. C. y Romero, N. J. G. (2024). *Transformación Educativa en la Era Digital: Integración y Futuro de las TIC en el Aprendizaje* (Vol. Editorial Internacional ALEMA). <https://editorialalema.org/libros/index.php/alema/article/view/33>
- Liu, F., Li, Z., Wang, B., Wu, J., Yang, J., Huang, J., Zhang, Y., et al. (2022). eRiskCom: an e-commerce risky community detection platform. *The VLDB Journal*, 31(5), 1085-1101. <https://doi.org/10.1007/s00778-021-00723-z>
- Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q. Z., Xiong, H., et al. (2021). A Comprehensive Survey on Graph Anomaly Detection With Deep Learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(12), 12012-12038. <https://doi.org/10.1109/TKDE.2021.3118815>
- Mendieta Lucas, L. M., Garzón Moreno, G. J., Enríquez Delgado, R. A. y Martínez Ángulo, M. A. (2025). Evolución e innovación digital en la educación superior como impulso para el fortalecimiento institucional: aplicación de inteligencia artificial en la gestión académica y administrativa para una toma de decisiones más eficiente y sostenible. *Reincisol*, 4(7), 2469-2492. [https://doi.org/10.59282/reincisol.V4\(7\)2469-2492](https://doi.org/10.59282/reincisol.V4(7)2469-2492)
- Mensah, G. A., Fuster, V., Murray, C. J. L., Roth, G. A., Mensah, G. A., Abate, Y. H., Abbasian, M., et al. (2023). Global Burden of Cardiovascular Diseases and Risks, 1990-2022. *Journal of the American College of Cardiology*, 82(25), 2350-2473. <https://doi.org/10.1016/j.jacc.2023.11.007>
- Mienye, I. D. y Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*, 10, 99129-99149. <https://doi.org/10.1109/ACCESS.2022.3207287>
- Mienye, I. D. y Sun, Y. (2023). A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection. *IEEE Access*, 11, 30628-30638. <https://doi.org/10.1109/ACCESS.2023.3262020>
- Morandín-Ahuerma, F. (2022). What is Artificial Intelligence? *International Journal of Research Publication and Reviews*, 3(12), 1947-1951. <https://doi.org/10.55248/gengpi.2022.31261>
- Mustafa, F., Almaududi Ausat, A. M. y Kraugusteeliana, K. (2024). The Role of Business Information Systems in Strategic Decision-Making: Implications for Innovation and Market Adaptation. *Jurnal Minfo Polgan*, 13(2), 1468-1475. <https://doi.org/10.33395/jmp.v13i2.14099>
- Naghavi, M., Ong, K. L., Aali, A., Ababneh, H. S., Abate, Y. H., Abbafati, C., Abbasgholizadeh, R., et al. (2024). Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*, 403(10440), 2100-2132. [https://doi.org/10.1016/S0140-6736\(24\)00367-2](https://doi.org/10.1016/S0140-6736(24)00367-2)
- Nelson Salgado, R. (2024). Automatización de tareas en sistemas de información: implementación de sistemas inteligentes para la automatización de tareas repetitivas y procesos rutinarios en entornos de sistemas de información. *Revista Científica FIPCAEC*, 9(1), 104-115. <https://doi.org/10.23857/fipcaec.v9i1.935>
- Ng, M., Gakidou, E., Lo, J., Abate, Y. H., Abbafati, C., Abbas, N., Abbasian, M., et al. (2025). Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. *The Lancet*, 405(10481), 813-838. [https://doi.org/10.1016/S0140-6736\(25\)00355-1](https://doi.org/10.1016/S0140-6736(25)00355-1)
- OECD. (2019). *The OECD Learning Compass 2030*. OECD Publishing. <https://www.oecd.org/en/data/tools/oecd-learning-compass-2030.html>
- OECD. (2023). *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem*. OECD Publishing. <https://doi.org/10.1787/c74f03de-en>
- Park, S.-M. y Kim, Y.-G. (2022). A Metaverse: Taxonomy, Components, Applications, and Open Challenges. *IEEE Access*, 10, 4209-4251. <https://doi.org/10.1109/ACCESS.2021.3140175>
- Quintero López, C., Gil Vera, V. D., Ángel Gómez, K., Valencia Ochoa, J., Restrepo Camargo, Y., Rivera Montoya, Y. y Herrera Vélez, A. (2024). Análisis de sentimientos sobre la educación remota en tiempos de pandemia: Latinoamérica. *Revista Palabra "palabra que obra"*, 24(1), 51-72. <https://doi.org/10.32997/2346-2884-vol.24-num.1-2024-4145>
- Rivera Aguilera, L. R., Rivera Aguilera, J. C. y Ramos Fandiño, G. P. (2024). Elementos a considerar en el diseño de sistemas de información para la gestión documental en instituciones públicas en México. En B. Cabral Vargas y J. M. Castillo Fonseca (Eds.), *Contexto y perspectiva del quehacer y desarrollo archivístico* (pp. 231-254). Universidad Nacional Autónoma de México. Instituto de Investigaciones Bibliotecológicas y de la Información. https://ru.iibi.unam.mx/jspui/handle/IIBI_UNAM/1056
- Romero, C. y Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>
- Sánchez Armas, M. y Delgado Bardales, J. M. (2020). Gestión Educativa en el desarrollo del aprendizaje en las Instituciones Educativas. *Ciencia Latina Revista Científica Multidisciplinar*, 4(2), 1819-1838. https://doi.org/10.37811/cl_rcm.v4i2.196
- Santos, C., Pedro, N. y Mattar, J. (2021). Competencia digital de profesores de educación superior:: análisis de factores académicos e institucionales. *Obra digital*, (21), 69-92. <https://doi.org/10.25029/od.2021.311.21>
- Sarker, I. H. (2022). AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science*, 3(2), 158. <https://doi.org/10.1007/s42979-022-01043-x>

- Sarker, I. H., Colman, A., Han, J. y Watters, P. (2021). *Context-Aware Machine Learning and Mobile Data Analytics: Automated Rule-based Services with Intelligent Decision-Making*. Springer. <https://doi.org/10.1007/978-3-030-88530-4>
- Schumacher, A. E., Kyu, H. H., Aali, A., Abbafati, C., Abbas, J., Abbasgholizadeh, R., Abbasi, M. A., et al. (2024). Global age-sex-specific mortality, life expectancy, and population estimates in 204 countries and territories and 811 subnational locations, 1950–2021, and the impact of the COVID-19 pandemic: a comprehensive demographic analysis for the Global Burden of Disease Study 2021. *The Lancet*, 403(10440), 1989-2056. [https://doi.org/10.1016/S0140-6736\(24\)00476-8](https://doi.org/10.1016/S0140-6736(24)00476-8)
- Steinmetz, J. D., Seeher, K. M., Schiess, N., Nichols, E., Cao, B., Servili, C., Cavallera, V., et al. (2024). Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet Neurology*, 23(4), 344-381. [https://doi.org/10.1016/S1474-4422\(24\)00038-3](https://doi.org/10.1016/S1474-4422(24)00038-3)
- Su, X., Xue, S., Liu, F., Wu, J., Yang, J., Zhou, C., Hu, W., et al. (2022). A Comprehensive Survey on Community Detection With Deep Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(4), 4682-4702. <https://doi.org/10.1109/TNNLS.2021.3137396>
- UNESCO. (2022). *Reimaginar juntos nuestros futuros: un nuevo contrato social para la educación*. UNESCO Publishing. <https://unesdoc.unesco.org/ark:/48223/pf00000381560>
- Van Gelder, I. C., Rienstra, M., Bunting, K. V., Casado-Arroyo, R., Caso, V., Crijns, H. J. G. M., De Potter, T. J. R., et al. (2024). 2024 ESC Guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): Developed by the task force for the management of atrial fibrillation of the European Society of Cardiology (ESC), with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Endorsed by the European Stroke Organisation (ESO). *European Heart Journal*, 45(36), 3314-3414. <https://doi.org/10.1093/eurheartj/ehae176>
- Wu, J., Lin, D., Chen, W., Zhang, L., Shen, X., Fu, D., Li, Y., et al. (2025). Research status and hotspots of oral frailty in older adults: a bibliometric analysis from 2013 to 2024. *Frontiers in Oral Health*, 6, 1533159. <https://doi.org/10.3389/froh.2025.1533159>
- Wu, J., Wang, Y., Shen, Z. y Liu, L. (2023). Adaptive client and communication optimizations in Federated Learning. *Information Systems*, 116, 102226. <https://doi.org/10.1016/j.is.2023.102226>
- Wu, J., Xu, Z., Huang, Q. y Cai, J. (2024). What motivates knowledge sharing? Evaluating the quality of answer contribution in online Q&A communities. *Data and Information Management*, 9(3), 100086. <https://doi.org/10.1016/j.dim.2024.100086>
- Yu, Z., Wu, J., Song, X., Fu, W. y Zhai, C. (2022). The Quantitative Research on Behavioral Intention towards 5G Rich Communication Services among University Students. *Systems*, 10(5), 136. <https://doi.org/10.3390/systems10050136>
- Zhou, S., Xu, H., Zheng, Z., Chen, J., Li, Z., Bu, J., Wu, J., et al. (2024). A Comprehensive Survey on Deep Clustering: Taxonomy, Challenges, and Future Directions. *ACM Computing Surveys*, 57(3), 1-38. <https://doi.org/10.1145/3689036>
- Zhu, D., Yin, H., Xu, Y., Wu, J., Zhang, B., Cheng, Y., Yin, Z., et al. (2023). A Survey of Advanced Information Fusion System: from Model-Driven to Knowledge-Enabled. *Data Science and Engineering*, 8(2), 85-97. <https://doi.org/10.1007/s41019-023-00209-8>