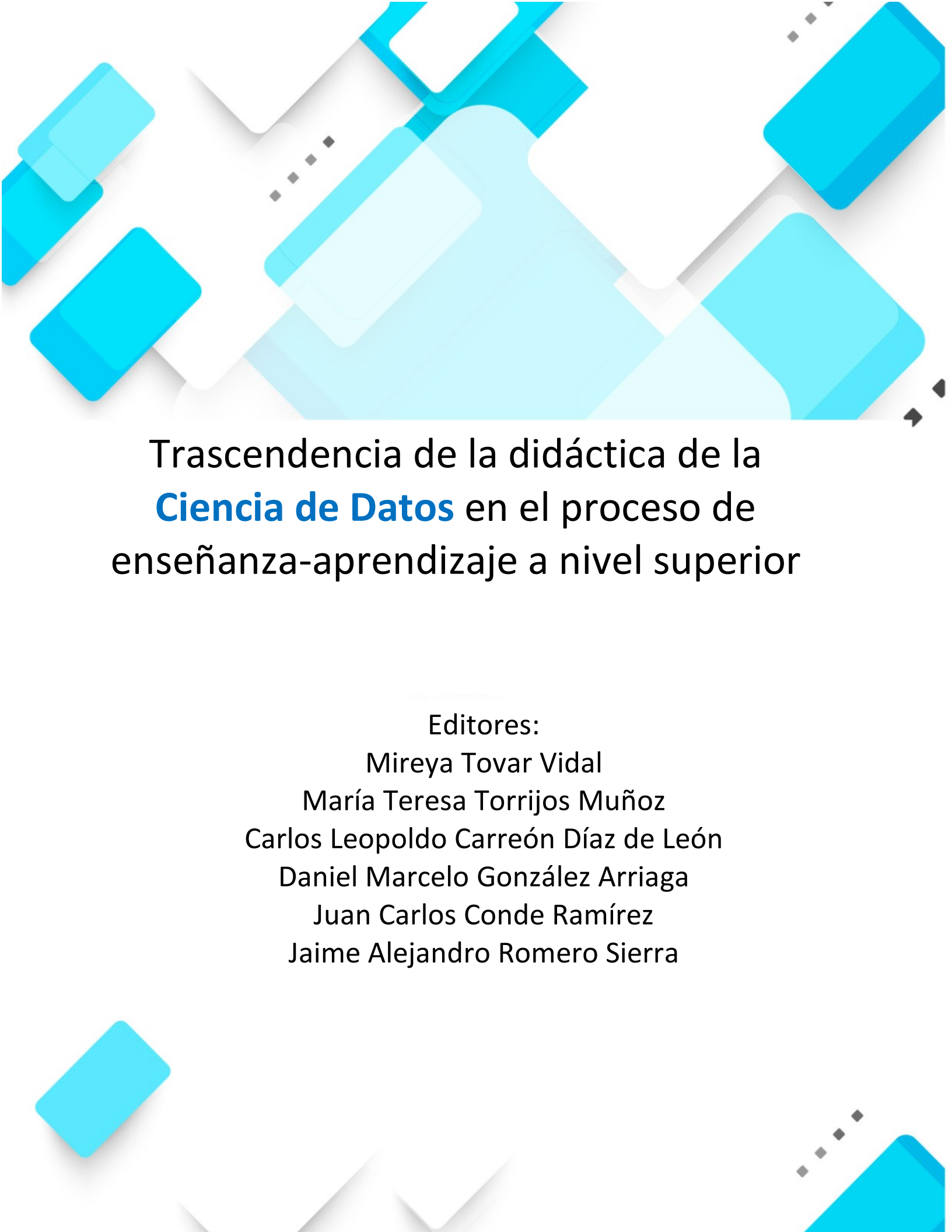




Trascendencia de la didáctica de la **Ciencia de Datos** en el proceso de enseñanza-aprendizaje a nivel superior

Editores:

Mireya Tovar Vidal

María Teresa Torrijos Muñoz

Carlos Leopoldo Carreón Díaz de León

Daniel Marcelo González Arriaga

Juan Carlos Conde Ramírez

Jaime Alejandro Romero Sierra

Trascendencia de la didáctica de la Ciencia de Datos en el proceso de enseñanza-aprendizaje a nivel superior

**Trascendencia de la didáctica de la Ciencia de
Datos en el proceso de enseñanza-aprendizaje a
nivel superior**

Mireya Tovar Vidal
María Teresa Torrijos Muñoz
Carlos Leopoldo Carreón Díaz de León
Daniel Marcelo González Arriaga
Juan Carlos Conde Ramírez
Jaime Alejandro Romero Sierra
Editores

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

RECTORA: MARÍA LILIA CEDILLO RAMÍREZ • SECRETARIO GENERAL: DAMIÁN HERNÁNDEZ MÉNDEZ • VICERRECTOR DE EXTENSIÓN Y DIFUSIÓN DE LA CULTURA: LUIS ANTONIO LUCIO VENEGAS • DIRECTOR GENERAL DE PUBLICACIONES: MARY KRISS PARRA GÓRRIZ

PRIMERA EDICIÓN **2025**
ISBN: 978-968-9744-03-0

D.R. © BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA
4 SUR 104, CENTRO HISTÓRICO, PUEBLA, PUE., CP 72000
TELÉFONO: 222 229 55 00
WWW.BUAP.MX

DIRECCIÓN GENERAL DE PUBLICACIONES
2 NORTE 1404, CENTRO HISTÓRICO, PUEBLA, PUE., CP 72000
TELÉFONO: 222 246 85 59
LIBROS.DGP@CORREO.BUAP.MX | WWW.PUBLICACIONES.BUAP.MX

DISEÑO DE PORTADA: ADOBE ILLUSTRATOR

HECHO EN MÉXICO
MADE IN MEXICO

Prólogo

En el presente libro titulado “*Trascendencia de la didáctica de la Ciencia de Datos en el proceso de enseñanza-aprendizaje a nivel superior*” se abordan diferentes áreas de investigación en didáctica en la Ciencia de Datos.

La obra incluye diez capítulos de investigación divididos en áreas de la educación orientadas a principios de la enseñanza-aprendizaje de la ciencia de datos.

Los capítulos que forman parte de esta obra fueron revisados mediante el sistema de doble par ciego y aprobados para su publicación por expertos en el área de conocimiento. De la misma manera la obra en su totalidad fue evaluada por dos expertos en el área de conocimiento. A continuación se menciona la aportación de cada uno de ellos.

En el Capítulo 1 se presenta un marco pedagógico para la enseñanza de la ciencia de datos a nivel universitario. Se analizan diversas estrategias de enseñanza-aprendizaje tales como el aprendizaje basado en proyectos, el aprendizaje activo y los interdisciplinarios como los métodos más apropiados para la formación de científicos de datos competentes.

En el Capítulo 2 se mencionan los desafíos que enfrentan los docentes en la formación de científicos de datos, y se destaca la necesidad de habilidades técnicas y pedagógicas en un campo multidisciplinario. Se enfoca en el marco TPACK (Conocimiento Tecnológico, Pedagógico y de Contenido) como herramienta para evaluar y mejorar las competencias docentes, y analiza métodos de enseñanza como el aprendizaje adaptativo, la educación basada en competencias, el aula invertida y la gamificación. El estudio remarca la importancia de los diferentes componentes del conocimiento, así como la formación continua de los docentes para adaptarse a los avances tecnológicos y garantizar una educación integral en ciencia de datos.

En el Capítulo 3 se destaca el papel de la estadística en la ciencia de datos junto a la programación para el desarrollo de modelos robustos y decisiones informadas. Se expone su evolución histórica, su papel en tareas como el análisis exploratorio, la validación de modelos y la inferencia, y cómo se complementa con la programación para resolver problemas reales. Además, aborda retos como el Big Data, la inferencia causal y la integración con inteligencia artificial, concluyendo que estadística y programación deben evolucionar juntas para una ciencia de datos sólida y efectiva.

En el Capítulo 4 se aborda cómo proteger la privacidad de los datos y, a la vez, facilitar su acceso para impulsar innovación y transparencia. Se analizan

tres tecnologías emergentes como el aprendizaje federado, privacidad diferencial y fideicomisos de datos, junto con casos y regulaciones en distintos sectores a nivel global y en México. Concluye que la combinación de tecnologías y marcos legales es esencial para lograr un equilibrio sostenible entre ambos objetivos.

En el Capítulo 5 se destaca la participación de estudiantes de ingeniería en ciencia de datos en datathons. También se muestran ejemplos en salud, finanzas y energía, y señala retos como el tiempo limitado, la coordinación del equipo y el manejo de datos complejos. Concluye que son una herramienta clave para el aprendizaje práctico, la innovación y la conexión entre estudiantes y la industria.

En el Capítulo 6 se analiza, mediante una revisión exploratoria, los factores que limitan el acceso equitativo a la ciencia de datos y sus implicaciones en el desarrollo educativo y científico, con énfasis en el contexto mexicano y latinoamericano. La investigación identifica como principales barreras la brecha digital, la carencia de infraestructura tecnológica, la falta de marcos regulatorios, el sesgo algorítmico, y las limitaciones idiomáticas. El trabajo plantea la necesidad de políticas educativas e institucionales para ampliar el acceso abierto y la equidad tecnológica.

En el Capítulo 7 se analiza los efectos de la democratización de la ciencia de datos en los roles profesionales, haciendo hincapié en desafíos como la calidad de los datos y los modelos, la escalabilidad, las mejores prácticas, las cuestiones éticas y la mitigación de sesgos. Concluye que los ingenieros en ciencia de datos siguen siendo esenciales como arquitectos de soluciones, educadores y guardianes de la confiabilidad.

En el Capítulo 8 se aborda la convergencia de tres áreas tecnológicas emergentes: arquitectura Brain-Computer Interface (BCI) basado en EEG, Internet de las Cosas (IoT) y ciencia de datos, enfocándose en aplicaciones que integran estas tecnologías para crear interfaces cerebro-computadora capaces de controlar dispositivos IoT. También discute la aplicación de técnicas de ciencia de datos para análisis descriptivo, diagnóstico y predictivo de las señales EEG, destacando el papel de algoritmos de aprendizaje automático en la interpretación de patrones cerebrales para generar comandos de control en tiempo real.

En el Capítulo 9 se propone un marco de referencia para orientar la acreditación de programas educativos de nivel superior en México afines a la Ciencia de Datos. Inicia dando una introducción describiendo lo que UNESCO y ANUIES definen como calidad en la Educación Superior; se mencionan los principales organismos de acreditación en México y los criterios para programas de Ciencias de Datos.

Por último, en el Capítulo 10 se presenta avances en el desarrollo de un sistema hipermedia adaptativo para un curso virtual de matemáticas financieras.

El sistema se desarrolló a partir de un diagnóstico de necesidades formativas de 66 estudiantes y se observó la experiencia de uso en un grupo focal de 10 estudiantes. Se reportan estadísticas descriptivas de las pruebas realizadas para el diagnóstico y las estrategias de aprendizaje elaboradas.

Finalmente, queremos agradecer a cada uno de los autores por su contribución, al comité revisor por su valiosa labor y cuidado en el proceso de revisión ciega, a la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla y a todos aquellos cuya participación contribuyó para la publicación de este libro.

Los editores,
Mireya Tovar Vidal
María Teresa Torrijos Muñoz
Carlos Leopoldo Carreón Díaz de León
Daniel Marcelo González Arriaga
Juan Carlos Conde Ramírez
Jaime Alejandro Romero Sierra

Índice general

Prólogo	IV
Capítulo 1. Pedagogías para la enseñanza del científico de dato	1
<i>Daniel Marcelo González Arriaga, Carlos Leopoldo Carreón Díaz de León, Juan Carlos Conde Ramírez, María Teresa Torrijos Muñoz</i>	
Capítulo 2. Desafíos de los docentes en la formación de los futuros científicos de datos	11
<i>Mireya Tovar Vidal, Jaime Alejandro Romero Sierra, Daniel Marcelo González Arriaga, Carlos Leopoldo Carreón Díaz De León</i>	
Capítulo 3. Más allá del código: La relevancia de la estadística en la ciencia de datos	16
<i>Héctor David Ramírez Hernández, Nelva Betzabel Espinoza Hernández, Roberto Contreras Juárez, Héctor Gabriel Ramírez Espinoza</i>	
Capítulo 4. Equilibrio entre privacidad y democratización	25
<i>Mauro Alberto López Muñoz, Berenice Rodríguez Pedroza, María Teresa Torrijos Muñoz, Carlos Armando Ríos Acevedo</i>	
Capítulo 5. Evaluación de la participación de estudiantes de ingeniería en ciencia de datos en datathons	36
<i>Luis Enrique Morales Aguilar</i>	
Capítulo 6. El impacto de la desdemocratización de la ciencia de datos en el desarrollo de la educación y la ciencia	44
<i>Berenice Rodríguez Pedroza, Mauro Alberto López Muñoz, María Teresa Torrijos Muñoz</i>	
Capítulo 7. El rol del Ingeniero en ciencia de datos en un entorno de creciente democratización de la disciplina	54
<i>Carlos Leopoldo Carreón Díaz De León, María Teresa Torrijos Muñoz, Mireya Tovar Vidal, Juan Carlos Conde Ramírez</i>	
Capítulo 8. Sistema de interfaz cerebro computadora basado en EEG en el contexto de Internet de las cosas y ciencia de datos – Una revisión .	61
<i>Regina Monserrat Avelino Farfán, Carlos Leopoldo Carreón Díaz de León, Daniel Marcelo González Arriaga</i>	
Capítulo 9. Marco de referencia para la acreditación de programas educativos de nivel superior en México afines a la Ciencia de Datos	73

*Juan Conde Ramírez, María Teresa Torrijos Muñoz, Mireya Tovar
Vidal, Jaime Alejandro Romero Sierra*

Capítulo 10. Desarrollo de un sistema hipermedia adaptativo para potenciar capacidades de aprendizaje en matemáticas financieras de nivel superior	86
<i>Eugenia Erica Vera Cervantes, Josefina Guerrero García, José Areli Carrera Román</i>	
Índice de autores	99
Compiladores	100
Revisores	101
Editores	102

Capítulo 1

Pedagogías para la enseñanza del científico de datos

Daniel Marcelo González Arriaga, Carlos Leopoldo Carreón Díaz de León, Juan Carlos Conde Ramírez, María Teresa Torrijos Muñoz

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, México

daniel.gonzalezarr@correo.buap.mx, carlos.carreon@correo.buap.mx,
carlos.conder@correo.buap.mx, teresa.torrijos@correo.buap.mx

Resumen. Formar a un científico de datos es un verdadero reto. No se limita a enseñar solo conceptos teóricos; es esencial tener una buena base técnica y un dominio práctico sólido. El pensamiento crítico también es clave para enfrentar problemas complejos. Este trabajo investiga las estrategias más efectivas para la enseñanza de la Ciencia de Datos en las universidades. Esto incluye métodos como el aprendizaje experiencial, el PBL y la colaboración activa, entre otros. También se destacan casos reales de éxito, como los "datathons" y laboratorios especializados. Sin embargo, la educación en este campo enfrenta grandes desafíos que deben ser atendidos. La falta de docentes experimentados y la dificultad para evaluar competencias por mencionar algunas. Este análisis sugiere maneras de unir la teoría con la práctica de manera efectiva.

Palabras Clave: Pedagogía, enseñanza, Ciencia de Datos, aprendizaje basado en proyectos, aprendizaje experiencial, educación superior, formación.

1 Introducción

La respuesta se encuentra en la Ciencia de Datos. Una disciplina que requiere personas expertas en la técnica, sin olvidar la habilidad de pensar críticamente y resolver problemas complejos (Wing, 2019).

Las universidades tienen un dilema aquí. Preparar a un científico de datos implica mucho más que solo enseñar matemáticas o programación de manera tradicional. La verdadera dificultad radica en enseñarles a pensar y aplicar lo que han aprendido, una tarea complicada para sistema educativo actual, que tiende a ser rígido y teórico, y que no se está logrando (Hicks & Irizarry, 2018).

Para abordar este desafío, muchas universidades experimentan con métodos de enseñanza innovadores. Desarrollan un aprendizaje que sea práctico y colaborativo en lugar de

depender de clases tradicionales. Utilizan proyectos basados en problemas (PBL), laboratorios de datos y "datathons", colocando a los estudiantes en situaciones reales (Saltz & Heckman, 2016; Goh & Sandars, 2020).

Aun con esto implementar estas pedagogías no es tarea sencilla. Las instituciones enfrentan grandes dificultades. La tecnología avanza muy rápido, faltan docentes especializados y es muy difícil evaluar, usando los métodos tradicionales las habilidades de los estudiantes (Baumer, 2015).

Por eso, en este artículo examinamos qué pedagogías funcionan mejor. Analizamos casos de éxito e identificamos los obstáculos más comunes. También proponemos soluciones prácticas para conectar la teoría con el mundo profesional (ver Figura 1).

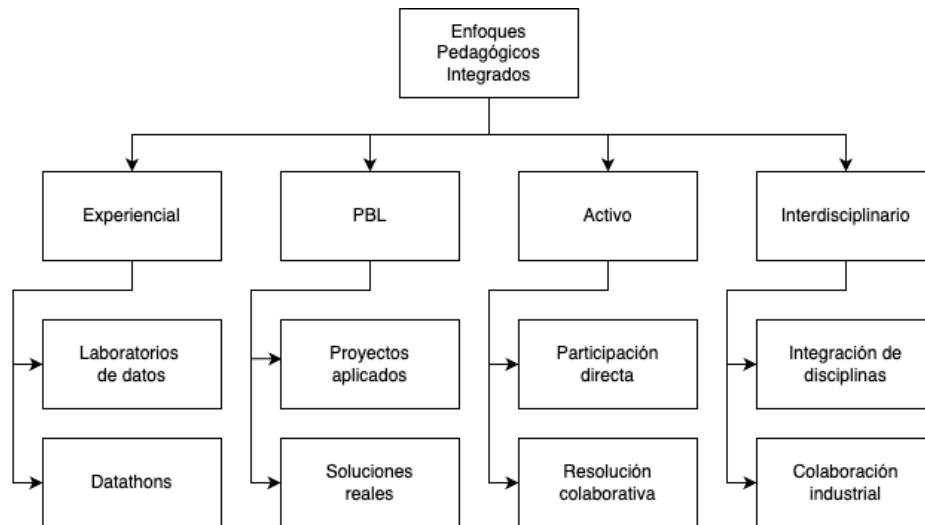


Figura 1 Diagrama conceptual sobre enfoques pedagógicos integrados para la enseñanza del científico de datos

2 Preliminares

La Ciencia de Datos es un área nueva en la educación. Integra estadística, informática y matemática aplicada. A veces también incluye el conocimiento del dominio donde se aplica. La formación de los científicos de datos debe ser completa. Necesitan conocimientos técnicos y pensamiento crítico. Deben saber resolver problemas complejos y comunicar sus resultados a múltiples públicos (Wing, 2019).

Hoy existe un consenso claro en la comunidad académica. Los métodos de enseñanza tradicionales ya no son suficientes. Las clases donde solo expone el profesor son muy limitadas para la Ciencia de Datos. Este campo demanda mucha práctica. Exige aplicar las habilidades en contextos apropiados (Hicks & Irizarry, 2018).

Otras pedagogías son mucho más efectivas. Por ejemplo, el aprendizaje obtenido por experiencia, el Aprendizaje Basado en Proyectos (ABP) y los enfoques interdisciplinarios. Estos métodos ayudan a adquirir habilidades que son aplicables en el mundo real (Saltz & Heckman, 2016).

El aprendizaje basado en la experiencia se explica a partir de la teoría de Kolb (1984). Su idea principal es simple: se aprende haciendo. La práctica y la reflexión continua sobre la práctica realizada son la base del conocimiento completo. Este enfoque resulta ser valioso en el área de la Ciencia de Datos. Esta disciplina requiere enfrentarnos a problemas reales y estar en un constante proceso de experimentación para así comprobar la teoría con la realidad.

El ABP desafía a los estudiantes a ensuciarse las manos. A través de proyectos que simulan desafíos reales, con esto no solo adquieren habilidades técnicas, sino que también se ven obligados a colaborar y a pensar de manera creativa para encontrar soluciones (Hmelo-Silver, 2004). Es el vínculo ideal entre el aula y el entorno laboral en la vida afuera de la universidad.

El enfoque es muy útil especialmente en disciplinas prácticas. La Ciencia de Datos es un buen ejemplo de esto. Esta área requiere interactuar continuamente con los problemas del mundo real. La experimentación continua es una parte integral del aprendizaje para los alumnos.

El Aprendizaje Basado en Proyectos (ABP) es diferente a lo anterior. Reta a los estudiantes a "ensuciarse las manos" y poner manos a la obra. Trabajan en proyectos que simulan desafíos del mundo real y no solo quedan en teoría. No solo aprenden la parte técnica, se enfrentan a problemas que a veces no considera la teoría y tiene que resolverlo sobre la marcha. También se ven forzados a colaborar, lo cual les ayuda a desarrollar la habilidad de trabajo en equipo que es muy importante. Deben pensar de forma creativa para encontrar soluciones (Hmelo-Silver, 2004). Este método es el puente perfecto entre el aula y el mundo laboral.

El aprendizaje activo invierte los roles normales. Usa herramientas como la clase invertida. El estudiante se convierte en el protagonista de su aprendizaje. El profesor, por su parte, actúa como un guía. El objetivo ya no es escuchar una lección. Se trata de construir el conocimiento en equipo. Esto se logra con debate e interacción constante (Prince, 2004).

Finalmente, están los enfoques interdisciplinarios. Estos entienden que la Ciencia de Datos es interdisciplinaria por naturaleza.

A pesar de sus ventajas, estas pedagogías necesitan un ambiente institucional adecuado. Se requieren docentes con formación específica para poder usarlas. También son necesarios criterios claros para medir el aprendizaje de forma objetiva. Las instituciones educativas necesitan políticas efectivas para resolver estos problemas. Así podrán maximizar los beneficios de estas metodologías (Ver Tabla 1 y Figura 1).

Tabla 1 Comparativa general de las pedagogías utilizadas en la enseñanza de Ciencia de Datos.

Pedagogía	Características clave	Ventajas	Desafíos	Ejemplos de aplicación
Aprendizaje Experiencial	Aprendizaje mediante prácticas en contextos reales.	Desarrollo rápido de habilidades prácticas y técnicas.	Requiere infraestructura especializada.	Laboratorios de datos, Datathons (competencias intensivas como Kaggle).
Aprendizaje Basado en Proyectos (PBL)	Resolución de problemas a través de proyectos reales o simulados.	Mejora del pensamiento crítico, trabajo colaborativo y aplicación práctica.	Evaluación compleja, requiere seguimiento continuo.	Proyectos integradores, análisis de casos reales con datos abiertos.
Aprendizaje Activo	Estudiantes como protagonistas activos del proceso de aprendizaje.	Incrementa compromiso, motivación y participación directa del estudiante.	Necesita preparación docente avanzada, diseño cuidadoso de actividades.	Clases invertidas (flipped classroom), actividades grupales interactivas.
Enfoques Interdisciplinarios	Integración de conocimientos y métodos de diferentes áreas.	Formación integral, desarrollo de habilidades transversales.	Coordinación curricular compleja, resistencia institucional al cambio.	Colaboraciones universidad-industria, cursos mixtos con expertos en distintas disciplinas.

3 Aprendizaje Experiencial en la Enseñanza del Científico de Datos

El aprendizaje experiencial, en términos generales, destaca la importancia de la obtención de conocimientos a través de la experiencia práctica, la cual debe ir acompañada de una reflexión crítica sobre la actividad llevada a cabo. En este caso, se sigue la teoría de Kolb (1984) que se basa en la reiteración de acción, observación, reflexión y conceptualización.

3.1. Aplicaciones prácticas en laboratorios de datos

Los Data Labs son espacios de trabajo donde los estudiantes se enfrentan a problemas reales donde pueden usar tecnología de punta como RStudio, Python y Hadoop. Allí pueden, por ejemplo, analizar series de tiempo con datos económicos o explorar grandes conjuntos de datos abiertos.

3.2. Datathons como herramienta experiencial

Por otro lado, los "datathons" son competencias intensivas que ponen a prueba a los estudiantes. En estos eventos, aplican sus conocimientos técnicos y analíticos bajo presión mientras trabajan en equipo. A nivel internacional, plataformas como Kaggle se han vuelto una vitrina excelente para que los nuevos talentos resuelvan desafíos prácticos.

4 Aprendizaje Basado en Proyectos (PBL)

El aprendizaje por proyectos sugiere que los estudiantes trabajen problemas complejos a través de grandes proyectos integradores, los cuales simulan situaciones del mundo real con la intención de aplicar conocimientos teóricos.

4.1. Diseño de proyectos integradores

La PBL permite que los estudiantes lleven a cabo proyectos que inician con la recolección y limpieza de datos hasta la presentación final de resultados. Este enfoque fomenta la adquisición de habilidades como trabajo en equipo, pensamiento crítico, comunicación y gestión efectiva del tiempo (Hmelo-Silver, 2004).

4.2. Vinculación con empresas y sectores profesionales

La vinculación que se realiza con la industria, en el caso con empresas tecnológicas o del sector financiero, hace que estas proporcionen conjuntos de datos y retos reales que tienen que ser resueltos por los estudiantes, de esta manera, se refuerza su perfil profesional.

5 Aprendizaje Activo para la Formación en Ciencia de Datos

El aprendizaje se hace horizontal, el protagonismo se transfiere a los estudiantes y, por consiguiente, cambia la manera en que se lleva a cabo el aprendizaje. Este se realiza a partir de actividades dinámicas, y los estudiantes tienen el control sobre el proceso.

5.1. Aula invertida

El aprendizaje en el aula permite que el estudiante controle la revisión de la teoría, el cual se trata en el aula mediante actividades físicas, así como el debate sobre los conceptos más complejos (Prince, 2004). En la Ciencia de Datos, esta metodología permite que en el tiempo de clase, se aproveche en la práctica de la programación y los debates sobre el uso de técnicas analíticas más complejas.

5.2. Resolución colaborativa de problemas

La resolución de problemas complejos de forma grupal permite a los estudiantes aprender unos de otros, consolidando habilidades interpersonales y técnicas. Es necesario destacar la importancia de definir estas actividades de forma equilibrada y con aportes permanentes de evaluación y de feedback constante por parte del docente.

6 Enfoques Interdisciplinarios en la Enseñanza de Ciencia de Datos

La falta de fronteras entre disciplinas permite abordar ciertos problemas de una Ciencia de Datos de manera más holística.

6.1. Colaboración académica interdepartamental

El desarrollo de interdisciplina de docentes de varias áreas (Informática, Estadística, Negocios, Ciencias Sociales) brinda una mayor riqueza y diversidad a los estudiantes (Goh & Sandars, 2020). Por ejemplo, en un proyecto de carácter interdisciplinario, la participación puede involucrar a estadísticos, ingenieros y a especialistas en ética, en especial en contextos de gran sensibilidad social o de elevado impacto

6.2. Alianzas estratégicas con la industria

Las alianzas estratégicas con empresas permiten a las universidades integrar aspectos prácticos y actualizados a su currículo académico. A través de seminarios, workshops, pasantías o proyectos patrocinados por empresas, los alumnos adquieren experiencia y conocimiento práctico en los aspectos de mayor relevancia y actualidad del mercado.

7 Desafíos en la Implementación de Pedagogías Innovadoras

A pesar de las considerables ventajas que ofrecen estas pedagogías, su implementación efectiva enfrenta numerosos problemas que requieren atención particular:

7.1. Rápida evolución tecnológica

Los programas académicos y docentes, así como las instituciones que los respaldan, enfrentan importantes problemas logísticos y organizacionales debido a la necesidad constante de actualización que implican las nuevas herramientas, técnicas y tecnologías que emergen de manera constante.

7.2. Formación especializada del profesorado

Por otro lado, la Enseñanza de la Ciencia de Datos con pedagogías de vanguardia demanda a los profesorados un proceso de formación más profundo, pues no solo deben poseer conocimientos técnicos, sino que necesitan habilidades pedagógicas creativas y herramientas didácticas pertinentes para poder dictar clases (Hicks & Irizarry, 2018).

7.3. Evaluación de competencias prácticas

Los métodos que se requieren para ofrecer retroalimentación a los estudiantes deben ser dinámicos, lo que, junto a la necesidad de evaluar las habilidades prácticas y las competencias que se engloban dentro de la evaluación formativa, con rúbrica, portafolios de trabajo, y evaluación práctica y continua resulta ser un desafío.

Estos problemas clave se sintetizan con claridad en la Figura 2, donde se muestran las pedagogías innovadoras a las que se les deben realizar cambios en su implementación e integración para atender las innovaciones curriculares que se requieren en la educación superior.

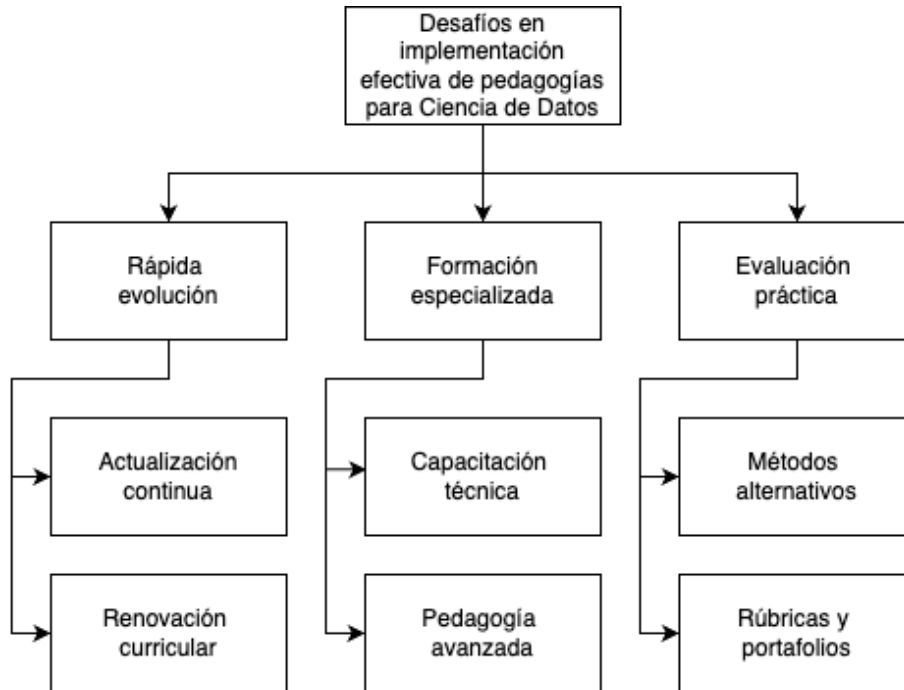


Figura 2 Diagrama de desafíos clave para la implementación efectiva de pedagogías innovadoras en la enseñanza de Ciencia de Datos.

8 Recomendaciones para la Implementación Efectiva de Pedagogías para Científicos de Datos

Con base en el análisis anterior, se establecen estrategias que atenderán de manera inmediata y efectiva las pedagogías referidas, en sus diversos niveles y en sus distintos contextos:

- Establecimiento de laboratorios especializados de datos con infraestructura tecnológica adecuada.
- Formación pedagógica especializada y continua para docentes.
- Vinculación formal con la industria mediante convenios estratégicos que permitan la actualización permanente del currículo.
- Desarrollo de metodologías claras y transparentes de evaluación práctica, utilizando rúbricas y portafolios digitales.

- Diseño curricular flexible que permita la integración efectiva de enfoques interdisciplinarios.

Implementar estas estrategias transformará la calidad de la educación superior. Así formaremos científicos de datos integrales, preparados para resolver los problemas más complejos que demanda el mundo profesional de hoy.

9 Conclusiones

La educación superior enfrenta hoy un gran desafío. Tiene que formar a los nuevos científicos de datos. Personas que sepan resolver los problemas complejos del mundo real.

Ya vimos que la teoría tradicional no es la solución. La respuesta está en usar pedagogías más dinámicas. Hablamos del aprendizaje experiencial, el ABP y el aprendizaje activo, por ejemplo.

¿Por qué funcionan estos métodos? Porque crean profesionales completos. No se trata solo de aprender la parte técnica. Los estudiantes también aprenden a pensar por sí mismos. Aprenden a comunicar sus ideas y a trabajar con otros.

Actividades como los "datathons" son perfectas para esto. Los laboratorios con datos reales también ayudan mucho. Conectan al estudiante con la industria. Mejoran sus oportunidades de trabajo. Y fuerzan a las universidades a modernizarse.

Claro, innovar no es un camino fácil. Pide mucho a las instituciones. Tienen que actualizar sus planes de estudio todo el tiempo. Deben preparar a sus profesores para nuevas formas de enseñar. Y necesitan encontrar una mejor manera de medir lo que un estudiante de verdad sabe hacer.

Para lograrlo, proponemos una hoja de ruta. Se centra en cuatro puntos clave:

- **Infraestructura:** Hay que invertir dinero en esto. Se necesitan laboratorios de datos modernos. Las plataformas tecnológicas deben estar al día.
- **Profesores:** Los docentes son la pieza central. Hay que capacitarlos constantemente. Deben dominar las nuevas herramientas y también las nuevas formas de enseñar.
- **Evaluación:** La forma de poner notas tiene que cambiar. Hay que usar métodos más reales y prácticos. Por ejemplo, revisar portafolios de proyectos o usar rúbricas claras.
- **Vinculación:** Conectarse con las empresas es obligatorio. Hay que crear alianzas fuertes con la industria. Es la única forma de que el plan de estudios no se quede obsoleto.

Seguir este plan es fundamental. Ayudará a las universidades a crear programas de Ciencia de Datos que sirvan. Y lo más importante: formará profesionales de alto nivel, listos para los retos del mercado laboral.

Referencias

- Baumer, B. S. (2015). A Data Science Course for Undergraduates: Thinking with Data. *The American Statistician*, 69(4), 334–342. <https://doi.org/10.1080/00031305.2015.1081105> (Artículo de revista)
- Goh, P. S., & Sandars, J. (2020). A vision of the use of technology in medical education after the COVID-19 pandemic. *MedEdPublish*, 9, 1–8. <https://doi.org/10.15694/mep.2020.000049.1> (Artículo de revista)
- Hicks, S. C., & Irizarry, R. A. (2018). A guide to teaching data science. *The American Statistician*, 72(4), 382–391. <https://doi.org/10.1080/00031305.2017.1356747> (Artículo de revista)
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16(3), 235–266. <https://doi.org/10.1023/B:EDPR.0000034022.16470.f3> (Artículo de revista)
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall. (Libro)
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3), 223–231. <https://doi.org/10.1002/j.2168-9830.2004.tb00809.x> (Artículo de revista)
- Saltz, J. S., & Heckman, R. (2016). Big data science education: A case study of a project-focused introductory course. *Themes in Science and Technology Education*, 8(2), 85–94. Recuperado de <https://www.learntechlib.org/p/174834/> (Artículo de revista)
- Wing, J. M. (2019). The data life cycle. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.e26845b4> (Artículo de revista)

Capítulo 2

Desafíos de los docentes en la formación de los futuros científicos de datos

Mireya Tovar Vidal, Jaime Alejandro Romero Sierra, Daniel Marcelo González Arriaga, Carlos Leopoldo Carreón Díaz De León

Facultad de Ciencias de la Computación, Benemerita Universidad Autonoma de Puebla, México

mireya.tovar@correo.buap.mx, jaime.romeros@correo.buap.mx,
daniel.gonzalezarr@correo.buap.mx, carlos.carreon@correo.buap.mx

Resumen. La Ciencia de Datos se caracteriza por ser un campo multidisciplinar que involucra el uso de tecnología, conocimientos en estadística, matemáticas, aprendizaje automático, etc., que se centra en extraer conocimiento y valor desde los datos. El proceso de enseñanza de las asignaturas en Ciencia de Datos requiere que el docente posea habilidades y competencias que le permitan el logro de objetivos de aprendizaje. También se presentan desafíos que el docente debe afrontar para el desarrollo de las habilidades que garanticen que el estudiante adquiere una formación integral, desarrolle competencias y lo prepare para los retos que surjan en el campo de la Ciencia de Datos. En este trabajo se presentan las habilidades, estrategias de enseñanza-aprendizaje y los retos que el docente puede llevar a cabo o enfrentar para facilitar el proceso de enseñanza-aprendizaje, en la impartición de las asignaturas de Ciencia de Datos.

Palabras clave: educación, ciencia de datos, métodos de enseñanza-aprendizaje.

1 Introducción

Como actualmente es bien sabido, la Ciencia de Datos es un nuevo campo de investigación que se centra en extraer conocimiento y valor desde los datos (Koby M., 2020). El crecimiento exponencial de los datos y de las tecnológicas para procesarlos ha generado un aumento en la demanda de científicos de datos. En Bureau of Labor Statistics (2023) se presenta una proyección de empleo para los científicos de datos del 36% del 2023 a 2033, lo que muestra su crecimiento y necesidad en la formación de estudiantes en ciencia de datos.

La educación en ciencia de datos es un campo académico multidisciplinario emergente que requiere enfoques pedagógicos para su impartición.

Un científico de datos requiere conocimientos en matemáticas, estadística, ciencias de la computación o áreas afines relacionadas a estos campos (Bureau of Labor Statistics, 2023). Debido a su transdisciplinariedad, la enseñanza en los programas de ciencia de datos plantean desafíos en los docentes (Theoren Msweli N., et. al., 2023). El docente requiere conocimientos teóricos y habilidades prácticas en ciencia de datos. Por lo que, es necesario desarrollar métodos pedagógicos, recursos didácticos, aplicaciones tecnológicas en el docente para impartir de manera integral la ciencia de datos de tal forma que contribuyan a la formación del estudiante y lo prepare para los retos que involucran soluciones basadas en datos.

El trabajo realizado se distribuye de la siguiente forma: En la sección 2 se presenta el marco TPACK; en la sección 3 algunos trabajos relacionados con la investigación; en la sección 4 se presentan algunas estrategias de enseñanza-aprendizaje; en la sección 5 métodos de enseñanza-aprendizaje; en la sección 6 las conclusiones del trabajo y finalmente las referencias consultadas.

2 TPACK en la educación en ciencia de datos

Uno de los métodos utilizados en el proceso de enseñanza-aprendizaje es el método TPACK (contenido tecnológico pedagógico). El cual consiste de los elementos: conocimiento del contenido (CK), conocimiento pedagógico (CP), conocimiento tecnológico (TK), conocimiento del contenido pedagógico (PCK), conocimiento del contenido tecnológico (TCK), conocimiento pedagógico tecnológico (TPK), conocimiento del contenido pedagógico tecnológico (TPACK, por sus siglas en inglés). A través de la tecnología y la pedagogía, el método permite al docente ofrecer material didáctico atractivo al estudiante. En la Tabla 1 se puede observar un resumen de los principales componentes del TPACK.

Tabla 1. Resumen de los componentes del TPACK (Mishra and Koehler (2008), Theoren Msweli, et. al. (2023))

Componente de conocimiento	Descripción
TK	Conocimiento sobre las tecnologías y herramientas.
CK	Los profesores deben poseer conocimientos y técnicas de ciencia de datos que incluyan la manipulación de datos, modelado estadístico, y aprendizaje automático.
PK	Los docentes deben tener experiencia en el uso de estrategias de enseñanza-aprendizaje que sean atractivas e interactivas y se adapten a los diferentes estilos de aprendizaje de los alumnos.

PCK	Es recomendable analizar la relación entre pedagogía, el contenido, las distintas disciplinas y como estás deben enseñarse con los métodos de enseñanza.
TCK	Consiste en comprender cómo el contenido y la tecnología impactan mutuamente.
TPK	Refiere al conocimiento de los alcances y limitaciones pedagógicas de las herramientas utilizadas.
TPACK	Los docentes deben utilizar la tecnología para mejorar el proceso de enseñanza-aprendizaje. Lo que incluye por ejemplo, el uso de plataformas, herramientas interactivas, material multimedia, recursos digitales, etc.

El método TPACK contribuye en el desarrollo de conocimientos, habilidades y capacidades profesionales (Theoren Msweli, et. al., 2023).

3 Trabajos relacionados

En la literatura existen trabajos que han aplicado el método TPACK con sólidos resultados. A continuación se presentan algunos de ellos.

El método TPACK ha sido ampliamente utilizado para medir las competencias desarrolladas en los docentes. En Yue, M., et. al., (2024) se presenta un estudio aplicado a 1664 docentes de distintos niveles académicos de enseñanza para determinar su preparación y actitudes hacia la enseñanza de la Inteligencia Artificial en el aula, a través de una encuesta que incluye los elementos del TPACK.

Theoren Msweli N., Mawala T., Twinomurinzi H. (2023) desarrollaron un instrumento basado en los siete constructos del TPACK, para determinar la capacidad de los instructores, en distintos niveles académicos, para enseñar ciencia de datos con apoyo de la tecnología. Además incorporaron al instrumento, el modelo de proceso CRISP-DM para proyectos en ciencia de datos. Los autores encontraron hallazgos interesantes como el análisis del estudio, por ejemplo, hay menos instructoras en el campo de la ciencia de datos y se tiene un nivel aceptable en los constructos de TPACK.

Leon Naranjo, J. R (2025) realiza un estudio para demostrar el efecto del modelo TPACK en las competencias digitales de los profesores de un centro educativo. La muestra estuvo formada por 70 docentes.

4 Habilidades docentes

Los docentes en ciencia de datos deben poseer habilidades técnicas y pedagógicas para enseñar esta disciplina.

5 Métodos educativos utilizados en la enseñanza de ciencia de datos

Uno de los desafíos más relevante en el proceso aprendizaje-enseñanza del docente en la impartición de asignaturas en Ciencia de Datos es desarrollar habilidades y competencias en el estudiante.

De acuerdo a Alcoba González (2013), los métodos de enseñanza son un conjunto de técnicas y actividades, que el docente usa para lograr un fin u objetivo educativo.

En la Tabla 2 se muestran algunos métodos de enseñanza y sus ventajas. A continuación, se mencionan algunos métodos que el docente emplea para la enseñanza de la ciencia de datos (Mejía, M. (2023).):

Aprendizaje adaptativo: es un aprendizaje que se puede adecuar en tiempo real al estudiante a través de herramientas como cuestionarios en línea, con la finalidad de reforzar debilidades detectadas a través de los mismos (Mejía, M. (2023)).

Educación basada en competencias: en este tipo el estudiante es el responsable del ritmo en el que aprende.

Aula invertida: es una estrategia autodidacta en la que el estudiante se empapa del conocimiento en casa y la práctica de lo aprendido la realiza en clase.

Tabla 2. Métodos de enseñanza y ventajas

Método	Ventajas
Aprendizaje adaptativo	Aprendizaje individual centrado en el estudiante. Las actividades están diseñadas para cada estudiante, lo que le permite avanzar más rápido.
Educación basada en competencias	La responsabilidad de educación es del estudiante, hasta que la logra pasa al siguiente conocimiento
Aula invertida	Técnica usada desde niveles básicos hasta posgrado. El estudiante se empapa del conocimiento en casa y la práctica de lo aprendido la realiza en clase

Gamificación	Se basa en el juego para potencializar la motivación, el interés, la concentración y el esfuerzo de los estudiantes. Los estudiantes se motivan, se divierten, se concentran, mejora el rendimiento académico, se incentiva el aprendizaje lógico.
--------------	---

6 Conclusiones

El estudio recomienda el desarrollo profesional de habilidades y competencias de manera continua de los docentes en ciencia de datos con el propósito que contribuyan en la formación de los futuros y actuales científicos de datos, esto debido a la creciente evolución y avance de la tecnología, los datos, las herramientas de ciencia de datos.

El método TPACK resulta útil para medir las capacidades de los docentes en ciencia de datos, que recomienda la inclusión de la tecnología en la impartición de cursos para lograr objetivos de aprendizaje, el desarrollo de habilidades y competencias en el estudiante.

Referencias

- Koby M. (2020). Data Science Education: Curriculum and pedagogy. In: Proceeding of the 2020 ACM Conference on International Computing Education Research, pp. 324 – 325. <https://doi.org/10.1145/3372782.3407110>
- Bureau of Labor Statistics, U.S. Department of Labor, Occupational Outlook Handbook, Data Scientists, at <https://www.bls.gov/ooh/math/data-scientists.htm> (visitado 23 de marzo 2025).
- Theoren Msweli N., Mawala T., Twinomurinzi H. (2023). Transdisciplinary teaching practices for data science education: A comprehensive framework for integrating disciplines. *Social Sciences & Humanities Open*, volume 8, issue 1. <https://doi.org/10.1016/j.ssaho.2023.100628>
- Yue, M., Jong, M.S.Y. and Ng, D. T. K. (2024). Understanding K-12 teachers' technological pedagogical content knowledge readiness and attitudes toward artificial intelligence education. *Education and Information Technologies*. 29:19505-19536. <https://doi.org/10.1007/s10639-024-12621-2>
- Mishra P. and Koehler M. J. (2008). Introducing technological pedagogical content knowledge. Annual Meeting of the American Education Research Association, New York City.
- Theoren Msweli N., Mawala T., Twinomurinzi H. (2023). Transdisciplinary teaching practices for data science education: A comprehensive framework for integrating disciplines. *Social Sciences & Humanities Open*, volume 8, issue 1. <https://doi.org/10.1016/j.ssaho.2023.100628>
- Leon Naranjo, J. R (2025). Implementación del modelo TPACK para fortalecer las competencias digitales de los docentes de una institución educativa. *IveCom*, 6(2), pp 1-9.
- Alcoba González, J. (2013). La clasificación de los métodos de enseñanza en educación superior. *Contextos Educativos. Revista de Educación*, (15), 93-106. <https://doi.org/10.18172/con.657>
- Mejía, M. (2023). Ciencia de Datos. Métodos educativos para su enseñanza. *Revista del Centro de Estudios Superiores en Educación*, innovus.mx

Capítulo 3

Más allá del código: La estadística en la ciencia de datos

Héctor David Ramírez Hernández, Nelva Betzabel Espinoza Hernández, Roberto Contreras Juárez, Héctor Gabriel Ramírez Espinoza

Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias de la Computación

hector.ramirezhe@correo.buap.mx, nelva.espinoza@correo.buap.mx,
roberto.contreras@correo.buap.mx, re202351956@alm.buap.mx

Resumen. La ciencia de datos se visualiza como una disciplina para la toma de decisiones en diferentes ámbitos. Esto ha sido gracias a los avances tecnológicos, el empleo de las redes sociales, así como de otros tipos de medios de comunicación. Generalmente, la ciencia de datos se relaciona principalmente con la programación y la aplicación de algoritmos, aunque sin perder de vista, la estadística se integra en ella, ofreciendo métodos y teoría para asegurar la confiabilidad, validez y la correcta interpretación de cualquier información basada en datos. El uso adecuado de la estadística permite obtener conclusiones tratando de disminuir el riesgo de sesgos o errores en las interpretaciones. En este trabajo se analiza la relación que existe entre la estadística y la programación para la ciencia de datos, mostrando que se complementan una con otra para resolver problemas en diversos contextos.

Palabras Clave: Ciencia de datos, programación, estadística.

1 Introducción

El volumen de datos crece exponencialmente y esta cantidad sigue aumentando día tras día. La información recolectada es vital para la toma de decisiones en casi todos los aspectos de la vida cotidiana. Bajo estas circunstancias, no es sorprenderse que en este contexto haya surgido la ciencia de datos como un conjunto de habilidades para manejar, usar y transformar estos datos en conocimiento que pueda ser utilizado para mejores predicciones, optimizar procesos y obtener nuevas oportunidades comerciales, así como de investigación.

Generalmente, se asume que la ciencia de datos es una disciplina dominada por la programación, donde la capacidad de escribir algoritmos en lenguajes como Python, R o SQL, entre otros, es necesaria tanto para el análisis como para la manipulación de datos, sin embargo, no se considera a un factor clave que es la estadística. Si no se cuenta con una base estadística sólida detrás de los modelos de aprendizaje automático predictivo más

complejos y modernos, entonces se pueden producir resultados inexactos y/o engañosos simplemente por no realizar el análisis estadístico correspondiente sobre los datos que se tienen.

A lo largo de este trabajo se resaltará el papel de la estadística en la ciencia de datos, destacando cómo su aplicación es crucial para desarrollar habilidades prácticas que resuelven problemas a través del análisis de datos en diferentes ámbitos, agregando un valor significativo cuando se conjunta con la programación.

La relación de la estadística con la ciencia de datos, la complementariedad entre la programación y la estadística, así como un enfoque sobre cómo lo deben de considerar las futuras generaciones.

Finalmente, se hará notar que la formación en ciencia de datos debe asegurar que la estadística pueda ser utilizada efectivamente para mejorar la toma de decisiones y prevenir errores debido a un mal análisis de datos. El objetivo es señalar la importancia de estudiar la ciencia de datos de manera holística, incluyendo la estadística como una herramienta fundamental de este campo.

2 Preliminares

El inicio del análisis estadístico data del siglo XVIII y se continuó con la incorporación de los conceptos clave de regresión y correlación formulados por Karl Pearson y Francis Galton en el siglo XIX (Pearson 1978). Estos permitieron el auge del análisis de datos, por lo cual matemáticos y científicos podían identificar patrones en grandes cantidades de datos.

El desarrollo del análisis de regresión y la estadística inferencial durante el siglo XX consolidó a la estadística. Los fundamentos incluyen el desarrollo de pruebas de hipótesis por Ronald Fisher a principios del siglo XX, y escritos sobre teoría de estimación de Jerzy Neyman y Egon Pearson en la década de 1930. Estos métodos son todavía importantes en la ciencia de datos moderna (Fisher, 1925). Andrey Kolmogorov también introdujo la axiomatización de la teoría de la probabilidad en 1933, que proporcionó una base para la toma de decisiones bajo incertidumbre. Estos avances fueron cruciales para darle sentido a los datos a través de la predicción.

En la década de 1950, con la tecnología informática moderna emergiendo y las bases de datos convirtiéndose en plataformas poderosas para que el trabajo académico en todas las áreas pudiera prosperar, la ciencia de datos fue reinventada. En ese momento, Edgar F. Codd, en su artículo "A Relational Model of Data for Large Shared Data Banks" publicado en la década de 1960, introdujo el modelo de base de datos relacional que era un modo efectivo de manipular grandes cantidades de datos. Esto llevó al desarrollo de sistemas para la acumulación y el almacenamiento de enormes volúmenes de información, derivando en la generación de algoritmos apropiados para recuperar y analizar datos de manera efectiva.

Durante las décadas de 1980 y 1990, el aprendizaje automático y la inteligencia artificial transformaron el análisis de datos. La automatización de la detección de patrones comenzó con algoritmos como las redes neuronales artificiales (Rumelhart, Hinton y Williams, 1986)

y los árboles de decisión (Quinlan, 1986). Estos avances impulsaron a la ciencia de datos al permitir la aplicación de modelos predictivos cada vez más complejos.

Con el tiempo, los datos han aumentado en volumen y se han desarrollado nuevos métodos para manejarlos y analizarlos desde principios del siglo XXI. William S. Cleveland acuñó el término "ciencia de datos" para referirse a una nueva disciplina cruzada que combinaba la informática con la estadística para abordar los desafíos del big data (Cleveland, 2001). Desde entonces, ha crecido muy rápidamente con métodos como el aprendizaje profundo y un análisis cuantitativo predictivo más completo.

Como ya hemos visto, la ciencia de datos es crucial para dominios diversos como la economía o la medicina. Empresas como Google, Facebook y Amazon han reinventado la toma de decisiones estratégicas a través del uso de datos (Provost y Fawcett 2013). Campos como la automatización de procesos industriales, el diagnóstico médico o la personalización de servicios han mejorado y se ha permitido su desarrollo, al menos parcialmente, porque la cantidad de datos que se pueden analizar los ha hecho factibles.

3 La importancia de la estadística en la ciencia de datos

Para crear, desarrollar, implementar y optimizar soluciones basadas en datos, un científico de datos necesita una amplia gama de habilidades técnicas, matemáticas y analíticas. Estas se pueden dividir en seis competencias esenciales que deben centrarse. Incluyen:

1. Programación y desarrollo de software,
2. Manejo y procesamiento de datos,
3. Estadísticas aplicadas y matemáticas,
4. Aprendizaje automático e inteligencia artificial,
5. Visualización y comunicación de datos, y
6. Ética y gobernanza de datos.

Un especialista en ciencia de datos con estas competencias puede realizar tareas a un nivel más sofisticado y abordar problemas complejos proporcionando soluciones razonables y eficientes basadas en datos.

Dado que la estadística proporciona una base matemática para la interpretación, modelado y análisis de datos, merece especial atención dentro de estas competencias. No se realizará un análisis de las otras competencias.

Usando técnicas estadísticas, se pueden ordenar enormes cantidades de datos y utilizarse para extraer información útil a todos los niveles, desde las llamadas estadísticas descriptivas hasta lo que se conoce como estadísticas inferenciales y conceptos más complejos.

El carácter estadístico de las conclusiones obtenidas es lo que proporciona a la disciplina su fundamento científico y preciso. Métodos como la recolección de datos, organización y representación descriptiva y visual, análisis de distribuciones e identificación y eliminación de valores atípicos ayudan a comprender la estructura de los datos antes de extrapolar

algunas tendencias subyacentes que podrían usarse para construir hipótesis probadas con rigor estadístico.

Una de las ramas más significativas de las estadísticas relevantes en la ciencia de datos es la estadística inferencial, orientada a predecir características de la población con el uso de muestras. Estas predicciones, además, deben estar respaldadas con evidencia estadística y procedimientos fácilmente repetibles. Es por eso por lo que se utilizan técnicas estadísticas como los intervalos de confianza y las pruebas de hipótesis (Casella y Berger, 2002), métodos que son aplicados en diversas disciplinas como la ingeniería de software, medicina, economía y otros campos.

La regresión estadística es otra herramienta en el análisis de datos. Las regresiones logística y lineal son modelos que se utilizan para categorizar datos y pronosticar tendencias (Hastie, Tibshirani y Friedman, 2009).

El aprendizaje automático también se basa en gran medida en la estadística. La optimización y evaluación de algoritmos como las máquinas de soporte vectorial, las redes neuronales y los árboles de decisión dependen de conceptos estadísticos (James et al., 2013).

La creación de modelos de aprendizaje automático puede dar lugar a sesgos y malentendidos si no se comprenden a fondo los conceptos estadísticos.

Basado en todo lo mencionado anteriormente podemos resumir que algunas aplicaciones clave de la estadística en ciencia de datos son:

- **Análisis exploratorio de datos (EDA):** Se identifican patrones, valores atípicos y correlaciones entre los datos antes de aplicar modelos avanzados (Tukey, 1977).
- **Validación de modelos:** Métodos como la validación cruzada y las métricas estadísticas (MSE, R^2 , AUC) permiten evaluar el desempeño de los modelos de machine learning (Hastie, Tibshirani y Friedman, 2009).
- **Inferencia y causalidad:** Diferenciar correlación de causalidad es clave para evitar interpretaciones erróneas de los datos (Pearl, 2009).
- **Reducción de dimensionalidad:** Son técnicas que permiten reducir la complejidad de los datos sin perder información relevante, como por ejemplo el Análisis de Componentes Principales (PCA). Esto es, el objetivo consiste en reducir a un número menor de datos si se considera un banco de datos con muchas variables, perdiendo la menor cantidad de información posible (Jolliffe y Cadima, 2016).

3.1 Caso práctico: la importancia de la estadística en la interpretación de resultados.

Un error común en ciencia de datos ocurre cuando se comparan resultados únicamente con medidas descriptivas sin aplicar un análisis estadístico adecuado. Por ejemplo, en la

evaluación de dos campañas de marketing digital sobre cuál es la más efectiva, para la Campaña A se obtuvo una tasa de conversión del 5% (50 ventas de 1000 visitas), mientras que para la Campaña B se alcanzó el 6% (30 ventas de 500 visitas).

Si se analiza solo el porcentaje, se concluiría erróneamente que la Campaña B es más efectiva. Sin embargo, al aplicar una prueba estadística de diferencia de proporciones (z-test), se obtiene un valor- $p \approx 0.416$, lo que indica que la diferencia no es significativa y que ambas campañas tienen un rendimiento similar.

Este ejemplo muestra que la programación permitió calcular rápidamente las tasas, pero fue la estadística la que permitió distinguir entre una variación aleatoria y una diferencia real. Sin este análisis, la empresa habría destinado recursos a una campaña sin evidencia sólida de que fuese mejor.

4 Estadística y Programación en ciencia de datos: Un enfoque complementario

Como se ha mencionado, la ciencia de datos debe combinar la programación, la estadística, así como el conocimiento del dominio para extraer valor de los datos (Provost y Fawcett, 2013). Su integración es ideal para desarrollar modelos y tomar decisiones informadas y confiables. Se destacará como se complementan estas dos disciplinas.

4.1 La programación en ciencia de datos

Se mencionó en la sección 3, que la programación es una de las seis habilidades principales que se desarrolla en la ciencia de datos, debido a que te permite modificar, analizar y visualizar grandes cantidades de datos.

Python y R son dos de los lenguajes más utilizados en la ciencia de datos debido a sus amplias bibliotecas especializadas, entre las que se incluyen NumPy, Pandas, SciPy y TensorFlow (McKinney, 2017).

La programación desempeña en la ciencia de datos funciones como:

- Extracción y limpieza de datos: La recopilación de datos en bruto desde diversas fuentes (APIs, bases de datos, archivos CSV) y su transformación en un formato estructurado y utilizable (Wickham y Grolemund, 2017).
- Automatización de procesos: La escritura de scripts permite ejecutar análisis repetitivos y entrenar modelos sin intervención manual.
- Implementación de modelos: Los algoritmos de machine learning requieren programación para ser entrenados, evaluados y desplegados en entornos productivos.
- Visualización de datos: Herramientas como Matplotlib, Seaborn y ggplot2 permiten crear gráficos para comunicar hallazgos de manera efectiva.

Se puede observar que la programación facilita la ejecución de los modelos, sin embargo, sin los conocimientos estadísticos suficientes y teniendo un exceso de confianza en los resultados de los algoritmos se dan interpretaciones inexactas.

4.2 Equilibrio entre estadística y programación

Si existe una armonía entre la programación y la estadística, la ciencia de datos será capaz de producir resultados precisos y útiles. Un análisis puramente estadístico sin herramientas computacionales puede resultar ineficaz debido al proceso de grandes volúmenes de datos y un modelo de ciencia de datos bien programado sin una validación estadística suficiente puede producir resultados inexactos (McKinney, 2017).

En la tabla 1 se presentan contribuciones de ambos campos a la ciencia de datos:

Aspecto	Programación	Estadística
Manipulación de datos	Uso de Pandas, SQL y herramientas de ETL	Identificación de sesgos en los datos
Modelado	Implementación de algoritmos de ML	Evaluación de supuestos y métricas
Interpretación	Aplicación de modelos predictivos	Validación de la significancia estadística
Visualización	Creación de dashboards y gráficos interactivos	Análisis de distribuciones y tendencias

Tabla 1 Contribución de la programación y la estadística a la ciencia de datos.

Se puede observar que la programación y la estadística se complementan entre sí. La estadística garantiza la validez y fiabilidad de los análisis, mientras que la programación permite automatizar procedimientos y ejecutar modelos de manera eficiente. Abordar con precisión los problemas del mundo real y adquirir habilidades prácticas en ciencia de datos requiere una estrategia integral que incorpore ambos campos.

5 Oportunidades y retos de la estadística en ciencia de datos

La relación entre la estadística y la ciencia de datos ha demostrado ser primordial para garantizar la precisión y fiabilidad de las decisiones basadas en datos. La estadística debe seguir estando a la vanguardia del desarrollo del campo, al tiempo que se adapta a los nuevos paradigmas computacionales, metodológicos y tecnológicos. A continuación, se muestran algunos temas de investigación, desde una perspectiva estadística, que los científicos de datos deben de considerar (He, 2020).

- Estadística y aprendizaje automático: el aprendizaje automático ha revolucionado el análisis de datos debido a que se crean modelos predictivos y prescriptivos cada vez más complejos. Dado que el aprendizaje automático se utiliza para tomar muchas decisiones, el sesgo algorítmico podría ser factor en la interpretación de los datos. Sin embargo, si el objetivo principal es la predicción, la interpretación pierde importancia. Es por ello que se debe asegurar que los resultados sean interpretables y que las conjeturas y sugerencias sean realistas. Así, el desempeño de la estadística en la ciencia de datos es importante, debido a la aplicación de técnicas de aprendizaje automático equitativas y comprensibles y toma en cuenta la ética y la reproducibilidad como criterios de evaluación.
- Ciencia de datos e inferencia causal: en la ciencia de datos es importante distinguir entre causalidad y correlación. Si no se consideran las bases de la estadística, los algoritmos de aprendizaje automático serían incapaces de inferir vínculos causales, a pesar de su capacidad para reconocer patrones y correlaciones.

Métodos como la inferencia causal basada en experimentos contrafactuales y los modelos causales gráficos son cada vez más populares (Pearl, 2009).

Es así como la estadística desempeña un papel más activo en la integración de técnicas de inferencia causal dentro del machine learning, ya que permite implementar modelos que no solo hacen predicciones, sino que también explican el porqué de sus decisiones (Peters, Janzing y Schölkopf, 2017).

- Estadísticas en el análisis de big data: A medida que crece la demanda de evaluar tipos de datos nuevos y complicados, como los datos falsos creados intencionadamente y la desinformación, la investigación estadística debe mantenerse al día. De manera similar, el big data ha creado una situación en la que los conjuntos de datos grandes y no estructurados pueden ser demasiado enormes para que los modelos estadísticos convencionales puedan manejarlos. Para manejar, limpiar y analizar eficazmente datos enormes, se requieren nuevos métodos y herramientas computacionales sofisticadas.

- Inferencia estadística escalable y distribuida basada en la computación en la nube: El análisis de macrodatos requiere técnicas computacionales sofisticadas e inferencia estadística escalable. El manejo de datos de alta dimensión requiere el uso de computación distribuida, aprendizaje estadístico distribuido y matrices dispersas. Aunque existen técnicas con sólidos fundamentos teóricos, resulta difícil escalarlas en entornos prácticos.

- Integración de la inteligencia artificial (IA): El rigor matemático que aporta la estadística al desarrollo de algoritmos de aprendizaje automático permite mejorar la selección de características, la validación de modelos y la reducción del sobreajuste.

Es así como la estadística ofrece contribuciones sólidas y prometedoras a la ciencia de datos debido a su integración con el análisis de big data, la inferencia causal y la inteligencia artificial.

La estadística tendrá que cambiar y evolucionar a medida que la tecnología avance y los conjuntos de datos crezcan en tamaño y complejidad, así como mantener su estrecha relación con la programación para seguir siendo clave en la solución de problemas complejos.

6 Conclusiones

Hemos abordado un panorama general de las contribuciones de la estadística y de cómo la tecnología mejora el procesamiento de datos para producir resultados confiables en la ciencia de datos.

Los avances en matemáticas, estadística e informática, así como, el crecimiento en el volumen de los datos y la creación de métodos analíticos, han dado lugar a los desarrollos actuales en la ciencia de datos.

A medida que la inteligencia artificial y el Big Data continúen evolucionando, la ciencia de datos continuará siendo pilar para el avance tecnológico y de la toma de decisiones basada en datos.

Se observó que la programación es una habilidad esencial para trabajar con grandes cantidades de datos, implementar modelos y automatizar operaciones, sin embargo, los conocimientos estadísticos se deben de combinar con ella para permitir modelos sólidos y análisis fiables.

La estadística tendrá muchos obstáculos que superar, pero también ofrece oportunidades para avanzar en la ciencia de datos, debido a que, uno de los principios fundamentales del análisis de datos sigue siendo su capacidad para descifrar, validar y mejorar los modelos matemáticos.

Es importante notar que se requieren de especialistas de ciencia de datos que se incorporen a la profesión estadística, así como a otras disciplinas que contribuyen a la ciencia de datos, especialmente la informática, la computación y los estudios éticos.

Referencias

Casella, G. and Berger, R.L. (2002) Statistical Inference. 2nd Edition, Duxbury Press, Pacific Grove.

Cleveland, W. S. (2001). Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics. International Statistical Review / Revue Internationale de Statistique, 69(1), 21–26. <https://doi.org/10.2307/1403527>

- Fisher RA. (1925) Theory of Statistical Estimation. Mathematical Proceedings of the Cambridge Philosophical Society. 22(5):700-725. doi:10.1017/S0305004100009580
- Hastie, T., Tibshirani, R., Friedman, J. (2009). Linear Methods for Regression. In: The Elements of Statistical Learning. Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-0-387-84858-7_3
- He, X. Lin, X. (2020) Challenges and Opportunities in Statistics and Data Science: Ten Research Areas. Harv Data Sci Rev. Summer; 2(3): 10.1162/99608f92.95388fcb. doi: 10.1162/99608f92.95388fcb. Epub 2020 Sep 30. PMID: 38116301; PMCID: PMC10730109.
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) An Introduction to Statistical Learning. Vol. 112, Springer, New York, 3-7. <https://doi.org/10.1007/978-1-4614-7138-7>
- Jolliffe, I., and Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A, 374(2065), 20150202.
- McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media.
- Pearl, J. (2009). Causality: Models, Reasoning, and Inference. Cambridge University Press.
- Pearson, K. (1978) The History of Statistics en the 17th and 18th Venturies, Edited by E. S. Pearson. New York: MacMillan.
- Peters, J., Janzing, D., Schölkopf, B. (2017). Elements of Causal Inference: Foundations and Learning Algorithms. The MIT Press.
- Provost, F., Fawcett, T. (2013) Data Science and its Relationship to Big Data and Data-Driven Decision Making. Vol. 1, Núm.1 Big Data. Edit. Mary Ann Liebert, Inc. <https://doi.org/10.1089/big.2013.1508>
- Quinlan, J.R. (1986) Induction of Decision Trees. Machine Learning, 1, 81-106. <http://dx.doi.org/10.1007/BF00116251>
- Rumelhart, D., Hinton, G. y Williams, R. (1986) Learning representations by back-propagating errors. Nature 323 , 533–536. <https://doi.org/10.1038/323533a0>
- Tukey, J. (1977). Exploratory Data Analysis. Reading Mass. Addison & Wesley.
- Wickham, H., and Grolemond, G. (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media.

Capítulo 4

Equilibrio Entre Privacidad y Democratización

Mauro Alberto López Muñoz¹, Berenice Rodríguez Pedroza², María Teresa Torrijos Muñoz³, Carlos Armando Ríos Acevedo⁴

¹ Benemérita Universidad Autónoma de Puebla,

² Benemérita Universidad Autónoma de Puebla,

³ Benemérita Universidad Autónoma de Puebla,

⁴ Benemérita Universidad Autónoma de Puebla,

¹mauroa.lopez@correo.buap.mx,

²berenice.rodriguezp@correo.buap.mx,

³teresa.torrijos@correo.buap.mx,

⁴carlos.rios@correo.buap.mx

Resumen. Este trabajo explora el equilibrio entre la privacidad y el acceso libre a la información. Se trata de un tema que ha ganado relevancia en una era en donde la ciencia de datos y la inteligencia artificial protagonizan la innovación tecnológica. El objetivo principal es realizar un estudio exploratorio para marcar un panorama general sobre las tecnologías, técnicas o métodos que buscan resolver los problemas que plantea el uso de datos y la confidencialidad. En primer lugar, se introducen los términos clave y su fuerte relación. Posteriormente, se mencionan técnicas y tecnologías emergentes que buscan equilibrar los 2 pilares y promover el desarrollo en diversas áreas. Luego se analizan diferentes contextos para dar a conocer el alcance del estudio propuesto. Finalmente se identifican desafíos y se plantean oportunidades para futuras investigaciones.

Palabras Clave: equilibrio, datos, privacidad, democratización, aprendizaje.

1 Introducción

La privacidad y la democratización de los datos son los pilares de este estudio, pero ¿qué es la privacidad? ¿Qué tan importante es en nuestras vidas?, y ¿qué papel juega la democratización para acceder a la información?

Kosinski y Forrest (2023) mencionan en su artículo para IBM que la privacidad en el contexto de la ciencia de datos se refiere a la protección de la información personal y sensible de los individuos. La definición implica medidas que aseguren que los datos no sean accesibles a terceros no autorizados y que se respeten los derechos de los sujetos de datos, Binns (2020).

Por otro lado, Awasthi y George (2020) explican que la democratización de datos se refiere a hacer que la información sea disponible y accesible para un mayor número de personas. Se trata de promover decisiones informadas y una mayor transparencia. De

esta manera se busca incentivar y empoderar a los usuarios finales permitiéndoles acceder y utilizar la información de manera eficiente y segura Wang et al. (2022).

Una vez establecida la concepción de los términos clave es natural plantear algunas interrogantes, ¿cómo surge el conflicto entre privacidad y democratización?, ¿qué medidas existen para lograr el equilibrio? y ¿por qué, hoy más que nunca es importante aplicar técnicas novedosas para proteger nuestra información?

Según Wasowicz (2025) surge un conflicto entre ambos conceptos cuando las medidas de protección de datos limitan el acceso a la información. En un principio el hecho de restringir acceso a ciertos datos puede resultar beneficioso, sin embargo, también puede obstaculizar la innovación y el desarrollo tecnológico. Por ejemplo, si se emplean técnicas de Differential Privacy se puede proteger la información sensible, pero también se puede limitar la utilidad de los datos para su análisis y toma de decisiones Sarathy et al. (2023). Se considera crucial este equilibrio para asegurar que se respeten los derechos individuales sin comprometer o limitar el potencial de los datos para generar valor social y económico Barocas et al. (2023).

Fu et al. (2024) aseguran que el conflicto entre la privacidad y la democratización en el uso y el acceso de datos comenzó a cobrar relevancia a medida que el análisis masivo, así como la recopilación de datos se volvieron más populares, con mayor énfasis en la última década. De este modo, desde alrededor de 2015, con la proliferación del Big Data y la rápida evolución de la inteligencia artificial, se ha intensificado el debate sobre cómo equilibrar la necesidad de proteger la privacidad de los individuos con la intención de democratizar el acceso a la información para impulsar el desarrollo, la innovación y la investigación Zhao et al. (2023).

El conflicto ha cobrado fuerza con la implementación de normativas. El Reglamento General de Protección de Datos (GDPR, por sus siglas en inglés) en Europa, introduce restricciones rigurosas en el tratamiento de la información personal. Si bien estas regulaciones resultan integrales para salvaguardar la privacidad, también pueden restringir el acceso a conjuntos de datos valiosos planteando así desafíos en la búsqueda de un equilibrio funcional y adecuado.

En México, la situación refleja desafíos similares. La Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP), en conjunto con la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPSO) plantean un marco normativo para asegurar la privacidad de los datos personales. Sin embargo, estas regulaciones también plantean retos para la democratización del acceso a la información. Olvera y Salmón plantean que es importante hacer énfasis en sectores como el académico y el gubernamental, donde el acceso a datos abiertos es fundamental para la investigación y la transparencia (2024; 2022).

Este trabajo presenta una revisión exploratoria en donde se seleccionaron fuentes de diversos artículos que abordan las técnicas que buscan equilibrar los 2 conceptos explicados al principio de esta sección. Se consideraron sitios oficiales de pioneros en la rama como Microsoft, Google y Apple, así como normativas oficiales en México y el mundo. Además, se complementó con páginas que abordan temas relevantes respecto a la investigación, así como contextos de interés.

2 Estrategias y tecnologías que buscan conciliar el conflicto

Existen diversas estrategias y tecnologías emergentes que buscan conciliar el conflicto entre privacidad y democratización. García Fernández (2024) destaca el Federated Learning, la Differential Privacy y los Data Trusts. Su objetivo es lograr el desarrollo tecnológico mientras protegen la información sensible.

En las siguientes secciones, se habla sobre cada una de estas estrategias, se exploran sus bases, aplicaciones y desafíos.

2.1 Federated Learning (FL)

El aprendizaje federado es el primer concepto clave. También conocido como Federated Learning o simplemente (FL), surge como respuesta a los desafíos de privacidad que representa el uso de las tecnologías actuales.

Fue introducido formalmente por Google. H. B. McMahan et al. (2017) explican que esta técnica ayuda a entrenar modelos de aprendizaje automático en múltiples dispositivos sin necesidad de centralizar los datos.

Lazaros et al. (2024) resalta su presencia en diversas aplicaciones y áreas como la salud, las finanzas y el Internet de las Cosas.

Se enfoca en el aprendizaje automático colaborativo. B. McMahan y Ramage (2017) explican que cada dispositivo descarga el modelo actual y lo mejora mediante el aprendizaje generado con los datos locales. Posteriormente resume los cambios en una pequeña actualización que se comparte en la nube. Finalmente se promedia con actualizaciones de otros usuarios para mejorar el modelo compartido y se reinicia el ciclo.

Así, este proceso asegura que los datos permanezcan privados en los dispositivos y no se compartan directamente.

A continuación, la Figura 1 presenta un diagrama que aproxima el funcionamiento de esta técnica.

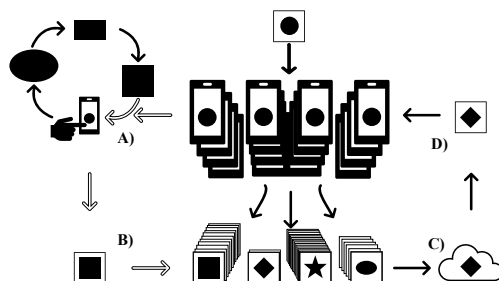


Figura 1 A) Cada cliente entrena el modelo localmente con sus propios datos. B) Los clientes envían las actualizaciones de su modelo (no los datos sin procesar) a un servidor central. C) El servidor agrega estas actualizaciones para mejorar el modelo global. D) El modelo actualizado se envía de vuelta a los clientes para su posterior entrenamiento.

Un ejemplo de implementación de la técnica es en los teclados Gboard de Google. Cada dispositivo recopila datos de texto de los usuarios y entrena localmente un modelo de predicción de texto. Las actualizaciones del modelo se comparten con el servidor central, mejorando así la precisión de las predicciones sin necesidad de enviar los datos textuales individuales Xu et al. (2023). Otro ejemplo se encuentra en el sector de la salud, donde los hospitales colaboran en el entrenamiento de modelos sin comprometer datos sensibles de los pacientes Weber (2024).

El aprendizaje federado busca conciliar el conflicto entre privacidad y democratización al permitir que múltiples entidades cooperen en el entrenamiento de modelos sin compartir datos sensibles. Así, Rieke (2019) afirma que esto facilita el acceso y la utilización de modelos de aprendizaje automático en diversas aplicaciones, promoviendo así la democratización del acceso a la tecnología.

2.2 Differential Privacy (DP)

El concepto de “privacidad diferencial” fue introducido por Dwork et al. (2006). Dwork junto a sus colegas desarrolló la técnica como respuesta al incremento de riesgos de privacidad relacionados con el análisis de grandes conjuntos de datos. Desde entonces, ha evolucionado para convertirse en un estándar de oro en la protección de la privacidad de los datos. Ha sido adoptado por empresas tecnológicas como Apple, Google y Microsoft para diversos usos y aplicaciones, desde el análisis de datos hasta la creación de modelos de aprendizaje automático.

La técnica es un marco matemático. Donde Dwork y Lei (2009) explica que “esta noción indica que cualquier resultado posible de un análisis debería ser casi igualmente probable, independientemente de si un individuo se excluye o no al conjunto de datos”. Es decir, que, al modificar una sola entrada en la base de datos, la salida de cualquier consulta debe ser estadísticamente similar, de modo que no se pueda identificar fácilmente qué entrada cambió.

Para ello, Li et al. (2023) argumentan que se introduce un nivel controlado de ruido aleatorio a los datos o a los resultados. Se mide mediante un parámetro llamado ϵ (épsilon). Dwork et al. (2006) muestran que si se considera un mecanismo de Laplace para la inyección de ruido el resultado perturbado es (1):

$$\mu(D) = f(D) + Y \quad (1)$$

Donde se establece ruido de Laplace con una desviación estándar $Y = \frac{S(f)}{\epsilon}$, con $S(f)$ definido como la sensibilidad, y $f(D)$ como el valor exacto que arroja la función f al aplicarse al conjunto de datos D , sin considerar el ruido.

Así, valores más bajos de ϵ proporcionan mayor privacidad al aumentar Y haciendo que predomine el ruido. Mientras tanto, valores más altos ofrecen mayor utilidad sacrificando la privacidad ya que Y disminuye y los resultados $\mu(D)$ son más precisos.

A continuación, en la Figura 2 se muestra un diagrama genérico que muestra el principio de funcionamiento de la privacidad diferencial.

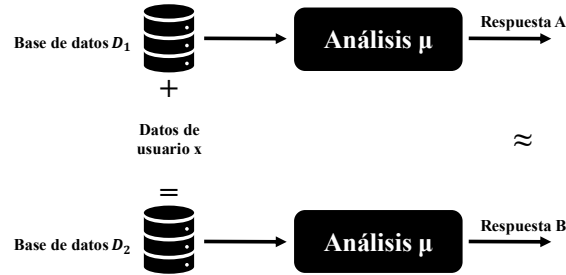


Figura 2 Diagrama que ejemplifica el funcionamiento de la privacidad diferencial. El “Análisis μ ” satisface privacidad diferencial si para todos D_1 y D_2 que difiere en los datos de un individuo, las respuestas A y B son indistinguibles.

Un ejemplo práctico de privacidad diferencial se presenta en “Learning with Privacy at Scale” publicado por Apple, donde se implementa para mejorar su teclado QuickType. Apple utiliza esta técnica para recopilar datos de uso de los teclados de los usuarios sin comprometer la privacidad (2017). Los registros agregados y anonimizados ayudan a mejorar la precisión y la funcionalidad del teclado. Otro ejemplo es su uso en el censo de los Estados Unidos, donde se aplicó para asegurar que la información proporcionada por los ciudadanos estuviera protegida contra la identificación individual, al tiempo que permitía la utilización de datos agregados para análisis demográficos Wood et al. (2020). Por otro lado, en el mismo año se examinó la aplicación de esta técnica en entornos clínicos. Se trató de una iniciativa conjunta entre Microsoft y Humana con el objetivo de facilitar el intercambio seguro de información médica (2020).

El conflicto entre privacidad y democratización busca ser conciliado al permitir el acceso y uso de grandes conjuntos sin comprometer la confidencialidad individual. Seeman y Susser (2024) aseguran que al introducir ruido controlado se permite que los investigadores y empresas utilicen datos de manera significativa y segura, promoviendo así la democratización del conocimiento y la innovación en armonía con la privacidad de los individuos.

2.3 Data Trusts

Los “fideicomisos de datos” (Data Trusts por su origen en inglés) surgen como una solución para abordar la creciente preocupación sobre la gobernanza y la gestión ética de los datos Zárate (2024). Aunque el concepto tiene raíces en prácticas fiduciarias tradicionales, Stachon et al. (2023) explican que la implementación de este enfoque en el ámbito de la gestión de datos adquirió mayor fuerza durante la última década, especialmente a partir de 2015.

El término refiere a una entidad fiduciaria que administra datos en nombre de un grupo de personas u organizaciones. Asegura que la información se use de manera ética, segura y en pro de los titulares Doyle (2024). Hardinges (2018) explica en su publicación “Defining a Data Trust” que una estructura fiduciaria significa que los fideicomisos de datos tienen un deber legal de actuar en el mejor interés de los titulares,

estableciendo normas claras sobre cómo se recopilan, almacenan, comparten y utilizan los datos

Un ejemplo notable de fideicomiso de datos es el Sidewalk Labs Urban Data Trust en Toronto. Dicho centro fue creado para gestionar los datos recopilados en un proyecto de ciudad inteligente Ghallab et al. (2018). Este centro tiene la responsabilidad de garantizar que los datos se utilicen de manera que beneficien a los residentes mientras se respeta su privacidad. Otro ejemplo es el MIDATA Cooperative en Suiza. En este centro los ciudadanos pueden almacenar sus datos de salud y decidir quién tiene acceso a ellos, promoviendo una gestión democrática y segura de la información personal (2019).

Se busca conciliar el conflicto entre privacidad y democratización al proporcionar una estructura que protege los derechos de confidencialidad de los individuos mientras se facilita el acceso y el uso de los datos.

El autor Montgomery (2021) explica en su artículo que estos centros permiten que los datos se usen de manera responsable y benéfica, además de promover la confianza y colaboración en su uso.

La Tabla 1 muestra las tecnologías y estrategias abordadas hasta el momento.

Tabla 1 Estrategias usadas para lograr el equilibrio entre privacidad y democratización.

Estrategia/ Tecnología	Descripción	Conciliación entre Privacidad y Democratización	Retos/Desafíos
Federated Learning	Técnica de aprendizaje automático donde el modelo se entrena en múltiples dispositivos sin centralizar los datos.	Protege la privacidad al no centralizar datos y facilita la colaboración y el acceso a modelos.	Coordinación compleja en múltiples modelos distribuidos. Carga computacional elevada y mayor consumo energético en equipos locales. Infraestructura costosa.
Differential Privacy	Técnica que introduce ruido estadístico a los resultados de consultas de bases de datos para proteger la privacidad.	Permite el acceso y uso de grandes conjuntos de datos sin comprometer la privacidad individual.	Necesita conocimientos avanzados en matemáticas y regulación de datos. La inyección de ruido puede ser compleja.
Data Trusts	Entidades fiduciarias que administran datos en nombre de un grupo, asegurando uso ético y seguro.	Protegen los derechos de privacidad de los individuos y facilitan el acceso y uso responsable de datos.	Viabilidad reducida en organizaciones pequeñas o con pocos recursos. Se debe trabajar en un marco

			legal global estándar y maduro.
--	--	--	---------------------------------

3 Casos y regulaciones en distintos contextos

Las regulaciones y casos específicos en los contextos académico, gubernamental y empresarial reflejan la tensión entre proteger la privacidad y promover la accesibilidad y el uso de los datos.

En las siguientes tres secciones se presentan ejemplos tanto a nivel mundial como nacional. Además, se muestra un panorama amplio sobre cómo se maneja este conflicto en distintos sectores.

3.1 Contexto académico

El acceso libre a los datos y su confidencialidad son fundamentales para facilitar la investigación en el sector académico. La normativa definida por el GDPR establece criterios rigurosos para el resguardo de la información personal (2020). Esto ha tenido una influencia importante en instituciones universitarias lo que ha llevado a tomar medidas para almacenar, recopilar y usar los datos.

Por otro lado, en México, el autor Ibarra Cadena (2025) escribe que el INAI regula la privacidad de los datos en el contexto académico. De esta manera busca garantizar que las sedes educativas cumplan con las normativas de protección de datos al mismo tiempo que promueven el acceso a la información.

3.2 Contexto gubernamental

Se han impulsado normativas para proteger la privacidad de los ciudadanos y permitir el acceso a la información pública.

En el norte de América, Bonta (2024) explica que la Ley de Privacidad del Consumidor de California es un ejemplo de legislación que equilibra la privacidad con el acceso a datos para la transparencia gubernamental.

Por otro lado, en el sector mexicano, la LGPDPSO establece los lineamientos para el tratamiento responsable de información personal por parte de entidades públicas (2017).

3.3 Contexto empresarial

A nivel mundial, Pelikan et al. (2023) analizan que empresas como Google y Apple utilizan tecnologías como la privacidad diferencial para proteger la información de los usuarios mientras mejoran sus servicios.

En México, la LFPDPPP regula cómo las empresas manejan los datos personales, promoviendo prácticas responsables y transparentes (2010). Además, se sedita equilibrar el acceso abierto a los datos empresariales con la protección de la privacidad de los consumidores mediante iniciativas como la creación de fideicomisos.

4 Vacíos y oportunidades para futuras investigaciones

La búsqueda de un equilibrio entre la privacidad y la democratización del acceso y uso de datos ha progresado significativamente en los últimos años. Sin embargo, aún existen vacíos importantes y oportunidades para futuras investigaciones tanto a nivel mundial como en México.

A pesar de que tecnologías como el aprendizaje federado y la privacidad diferencial han impactado a nivel mundial, aún es necesario indagar cómo estas y otras técnicas pueden llegar a más contextos y sectores Ni y Zhou (2025)

Mientras tanto en México Chávez (2018) afirma que el país tiene carencias en la evaluación de implementación y efectividad de las regulaciones abordadas en secciones anteriores. Además, es importante estudiar con mayor detenimiento el impacto de las políticas de privacidad en los sectores educativo, gubernamental y empresarial.

La creación de estructuras de gobernanza de datos inclusivos y equitativos es otra área de oportunidad.

Mientras tanto, Riya et al. (2023) dicen que es muy importante investigar las implicaciones éticas y sociales de las tecnologías que van apareciendo. Además, es útil analizar cómo las percepciones públicas de privacidad y gestión de datos pueden influir en la adopción de estas nuevas tendencias.

La poca confianza en las instituciones públicas en México es un desafío real. Por ello, resulta importante trabajar en estudios sobre cómo fortalecer la transparencia y la rendición de cuentas en el manejo de la información.

5 Conclusiones

El estudio realizado logró aterrizar la definición de conceptos clave y las relaciones entre ellos. Se abordaron tendencias tecnológicas que buscan promover el desarrollo y la investigación sin sacrificar la privacidad de la información.

Al profundizar en cada técnica mencionada se observó que, a pesar de contar con grandes avances, aún queda un largo camino para su adopción e implementación en países en vías de desarrollo. Es importante mencionar que, aunque los métodos presentados pueden ser una solución también pueden representar grandes desafíos por el costo computacional que implica su uso.

Tras el análisis de casos en diferentes contextos se concluye que es fundamental crear, revisar y actualizar marcos regulatorios que permitan explotar el potencial de la información al mismo tiempo que mantienen la integridad y confidencialidad de los datos. Además, se resalta la premisa de invertir en infraestructura, investigación y desarrollo para lograr adaptarnos a los problemas y desafíos que se presentan en la actualidad.

Finalmente se identificaron vacíos y se plantearon oportunidades para futuras investigaciones o incluso implementaciones de las estrategias mencionadas a lo largo del estudio realizado.

Referencias

- Awasthi, P., & George, J. J. (2020). A case for Data Democratization. *AMCIS 2020 Proceedings*, 23.
- Barocas, S., Hardt, M., & Narayanan, A. (2023). Our outlook: limitations and opportunities. In *Fairness and Machine Learning Limitations and Opportunities* (pp. 1–22). MIT Press.
- Binns, R. (2020). On the apparent conflict between individual and group fairness. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 514–524. <https://doi.org/10.1145/3351095.3372864>
- Bonta, R. (2024, March 13). *California Consumer Privacy Act (CCPA)*. State of California Department of Justice. <https://www.oag.ca.gov/privacy/ccpa>
- Chavez, C. A. (2018). *Data Privacy Compliance under the Mexican Law and the GDPR*.
- Differential Privacy Team. (2017, December 6). *Learning with Privacy at Scale*. Machine Learning Research.
- Doyle, J. (2024, December 11). *What is a Data Trust? The complete guide for organizations, regulators and manufacturers*. Creme Global. <https://www.cremeglobal.com/what-is-a-data-trust-the-complete-guide-for-organizations-regulators-and-manufacturers/>
- Dwork, C., & Lei, J. (2009). Differential privacy and robust statistics. *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, 371–380. <https://doi.org/10.1145/1536414.1536466>
- Dwork, C., Mcsherry, F., Nissim, K., & Smith, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis. *Springer*, 265–284.
- Fu, J., Hong, Y., Ling, X., Wang, L., Ran, X., Sun, Z., Wang, W. H., Chen, Z., & Cao, Y. (2024). Differentially Private Federated Learning: A Systematic Review. *ACM*. <http://arxiv.org/abs/2405.08299>
- García Fernández, C. (2024, March 6). *Ética en Data Science: Desafíos y consideraciones clave*. Open Webinars. <https://openwebinars.net/blog/etica-en-data-science/>
- Ghallab, M., Delacroix, S., Penn, J., Colclough, C., Santow, E., Howell, J., McDonald, S., Porcaro, K., Klein, H. B., Mullin, S., Viljoen, S., & Lawrence, N. (2018). *Data Trusts A new tool for data governance*.
- Hardinges, J. (2018, July 10). *What is a data trust?* Open Data Trust. <https://theodi.org/insights/explainers/what-is-a-data-trust/>
- How differential privacy enhances Microsoft’s privacy and security tools: SmartNoise Early Adopter Acceleration Program Launched. (2020). *Microsoft on the Issues*. <https://blogs.microsoft.com/on-the-issues/2020/12/10/differential-privacy-smartnoise-early-adopter-acceleration-program/>
- Ibarra Cadena, B. L. (2025). *Resoluciones en materia de Acceso a la Información y Protección de Datos Personales*. INAI. https://home.inai.org.mx/?page_id=1641&op=5
- Kosinski, M., & Forrest, A. (2023, December 19). *¿Qué es la protección de los datos?* IBM. <https://www.ibm.com/es-es/topics/data-privacy>
- Lazaros, K., Koumadorakis, D. E., Vrahatis, A. G., & Kotsiantis, S. (2024). Federated Learning: Navigating the Landscape of Collaborative Intelligence. In *Electronics (Switzerland)* (Vol. 13, Issue 23). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/electronics13234744>

- Ley Federal de Protección de Datos Personales en Posesión de los Particulares*. (2010). <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>
- Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados*. (2017). <https://www.diputados.gob.mx/LeyesBiblio/pdf/LGPDPPSO.pdf>
- Li, Y., Liu, Y., Li, B., Wang, W., & Liu, N. (2023). Towards practical differential privacy in data analysis: Understanding the effect of epsilon on utility in private ERM. *Computers and Security*, 128. <https://doi.org/10.1016/j.cose.2023.103147>
- McMahan, B., & Ramage, D. (2017). Federated Learning: Collaborative Machine Learning without Centralized Training Data. *Google Research*. <https://research.google/blog/federated-learning-collaborative-machine-learning-without-centralized-training-data/>
- McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Aguera, B. (2017). *Communication-Efficient Learning of Deep Networks from Decentralized Data*.
- MIDATA Genossenschaft. (2019). https://www.midata.coop/wp-content/uploads/2019/08/MIDATA_Statuten_20190626_EN.pdf
- Montgomery, J. (2021, March 1). *International perspectives on data institutions: lessons for data trust*. Data Trusts Initiative. <https://datatrusts.uk/2021/03/01/international-perspectives-on-data-institutions-lessons-for-data-trusts.html>
- Ni, Z., & Zhou, Q. (2025). Differential Privacy in Federated Learning: An Evolutionary Game Analysis. *Applied Sciences*, 15(6), 2914. <https://doi.org/10.3390/app15062914>
- Olvera Arellano, A. G. (2024, January 26). *La importancia de la protección de datos personales en la era digital*. Revista Digital Del Instituto de Información Estadística y Geográfica. <https://iieg.gob.mx/strategos/la-importancia-de-la-proteccion-de-datos-personales-en-la-era-digital/>
- Pelikan, M., Azam, S. S., Feldman, V., Silovsky, J., “Honza,” Talwar, K., & Likhomanenko, T. (2023). *Federated Learning with Differential Privacy for End-to-End Speech Recognition*. <http://arxiv.org/abs/2310.00098>
- Reglamento general de protección de datos (RGPD)*. (2020, June). EUR-Lex. <https://eur-lex.europa.eu/ES/legal-content/summary/general-data-protection-regulation-gdpr.html>
- Rieke, N. (2019, October 13). *What Is Federated Learning?* Nvidia. <https://blogs.nvidia.com/blog/what-is-federated-learning/>
- Riya, M., Student, C., & LI, B. A. B. (2023). Data Privacy and the Legal Implications of Emerging Technologies. *JETIR*, 10. www.jetir.org
- Salmón, E. (2022, September 13). *Democratización de la información y su importancia para la protección y promoción de los DDHH: retos y avances en la era digital*. <https://idehpucp.pucp.edu.pe/boletin-eventos/democratizacion-de-la-informacion-y-su-importancia-para-la-proteccion-y-promocion-de-los-ddhh-retos-y-avances-en-la-era-digital-26981/>
- Sarathy, J., Song, S., Haque, A., Schlatter, T., & Vadhan, S. (2023, April 19). Don't Look at the Data! How Differential Privacy Reconfigures the Practices of Data Science. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3544548.3580791>
- Seeman, J., & Susser, D. (2024). Between Privacy and Utility: On Differential Privacy in Theory and Practice. *ACM Journal on Responsible Computing*, 1(1), 1–18. <https://doi.org/10.1145/3626494>

- Stachon, M., Möller, F., Guggenberger, T., Tomczyk, M., Braunschweig, T. U., & Henning, J.-L. (2023). Understanding Data Trusts. *Thirty-First European Conference on Information Systems*.
- Wang, Y., Blobel, B., & Yang, B. (2022). Reinforcing Health Data Sharing through Data Democratization. *Journal of Personalized Medicine*, 12(9). <https://doi.org/10.3390/jpm12091380>
- Wasowicz, P. (2025). *Democratización de datos - Privacidad en arquitecturas de streaming*. Mima.com. <https://www.mimacom.com/es/blog/privacidad-de-datos-vs-democracia-de-datos>
- Weber, M. (2024). *An In-Depth Guide to Federated Learning Concepts and Techniques*. Kinda Technical. <https://kindatechnical.com/federated-learning-guide/history-and-evolution-of-federated-learning.html>
- Wood, A., Altman, M., Bembeneck, A., Bun, M., Gaboardi, M., Honaker, J., Nissim, K., O'brien, D. R., Steinke, T., Vadhan, S., Abowd, J., Bradner, S., Dwork, C., Garfinkel, S., Gooden, C., Hurley, D., Kalmar, R., Kellaris, G., Muise, D., ... Washington, M. (2020). *Differential Privacy: A Primer for a Non-Technical Audience*. <http://privacytools.seas.harvard.edu>
- Xu, Z., Zhang, Y., Andrew, G., Choquette, C., Kairouz, P., McMahan, B., Rosenstock, J., & Zhang, Y. (2023). Federated Learning of Gboard Language Models with Differential Privacy. *Google Research*. <https://research.google/pubs/federated-learning-of-gboard-language-models-with-differential-privacy/>
- Zárate, M. (2024, February 24). *Fideicomiso de Datos: ¿Cómo blindar la información personal?* Escala Legal. <https://escalalegal.com/2024/02/24/fideicomiso-de-datos-como-blindar-la-informacion-personal/>
- Zhao, R., Goel, N., Agrawal, N., Zhao, J., Stein, J., Verborgh, R., Binns, R., Berners-Lee, T., & Shadbolt, N. (2023). *Libertas: Privacy-Preserving Computation for Decentralised Personal Data Stores*.

Capítulo 5

Evaluación de la participación de estudiantes de ingeniería en ciencia de datos en datathons

Luis Enrique Morales Aguilar

Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias de la Computación

`luis.moralesagu@correo.buap.mx`

Resumen. En este capítulo se analizó la participación del alumnado en eventos llamados Datathons, como es que se clasifican y más importante, como es que la participación de los alumnos influye de manera positiva en su formación profesional mediante el desarrollo y fortalecimiento de habilidades tanto técnicas, como las llamadas “Softskills”. Además se proporcionan algunos consejos para el estudiante con el fin de tener una mejor experiencia.

Palabras Clave: Datathons, Ciencia de Datos, Trabajo En Equipo, Networking.

1 Introducción

Las nuevas tecnologías han permitido el desarrollo a gran escala de la ciencia de datos que hoy en día se considera un pilar fundamental del mundo de la computación. El análisis de grandes cantidades de datos que hasta hace unas décadas podría tomar horas o hasta días, en la actualidad toman muchísimo menos tiempo, lo cual permite obtener información clave que se usa como una herramienta en la toma de decisiones. Como plantea Lemus-Delgado y Navarro (2020) el análisis de datos puede ser empleado en cualquier campo de la sociedad, por ejemplo; salud y economía, nos permite mediante modelos complejos predecir el comportamiento de enfermedades o de acciones en la bolsa, o con modelos más simples para clasificar objetos o realizar una regresión lineal.

García et al. (2018) definen a la ciencia de datos como la capacidad de transformar inimaginables cantidades de datos, por decirlo de alguna manera “crudos”, en información útil para el desarrollo de ideas, para fundamentar conceptos o como un soporte de decisiones que tienen un impacto profundo en la sociedad. Todo esto se lleva a cabo mediante la fusión de la estadística, cientos de fuentes de datos y herramientas computacionales que permiten una correcta interpretación de datos. Ahora con la llegada de la “Code generation AI” o “Copilots”, es posible que cualquier persona pueda generar scripts y comience con el análisis de grandes datasets sin la necesidad de ser un experto en esta ciencia, lo que trae consigo que cada vez haya mayor competencia.

Con el surgimiento de eventos como congresos y hackatones en el ámbito de las ciencias computacionales y gracias a la amplia expansión de la ciencia de datos además de fomentar la colaboración, surgen los Datathons, una rama de los hackatones que se centra en el análisis de masivas cantidades de datos con el fin de dar solución a problemas de la sociedad. El instituto internacional de estadística fundó en 1977, la asociación internacional de computación y estadística (IASC, 1977), a su vez, la asociación fue quien comenzó con la idea de los congresos centrados específicamente en el análisis de datos y por ello considerados por muchos como los pioneros de lo Datathons, el COMPSTAT desde 1998 y DSSV desde 2017, fomenta en los participantes el uso de la creatividad y el uso de habilidades analíticas en un ambiente competitivo, en donde los competidores son sometidos a trabajar bajo presión y muchas veces en equipos que no se conocían hasta dicha fecha. (Data Analysis Competition | IASC-ISI, s. f.).

En las últimas décadas los datathons han sido tópicos de interés, Anslow et al. (2016) describe a los datathons como eventos en los cuales, los participantes formados en equipos, muchas veces formados por elementos multidisciplinarios, ponen a prueba sus capacidades para, mediante el análisis de datos, obtener información en tiempo récord, la cual permita brindar una solución sustentada a un problema específico. Recientemente (2021), el ministerio de salud de Israel organizó un datathon con el fin de dictar recomendaciones y obtener modelos descriptivos de la enfermedad y en tan solo 48 horas los equipos interdisciplinarios completaron sus objetivos.

DSSV es una conferencia anual que tiene como objetivo reunir a investigadores, profesionales y principiantes del área de la ciencia de datos que estén interesados en la interacción conjunta de la informática, visualización y la estadística, a formar puentes de conocimiento con distintas disciplinas. El foro permite a los participantes discutir los recientes avances e ideas emergentes que permitan el desarrollo de la ciencia de datos, análisis y visualización de big data (Home, s. f.).

La asociación internacional de estadística y computación, en su sección regional asiática (ASC-ARS), tiene la misión de organizar seminarios, conferencias y reuniones internacionales del ámbito de la estadística computacional, además de publicar libros y periódicos del tema. En 2024 como cede la Universidad Nacional de Taiwán, se llevó a cabo la Interim Conference 2024, en la cual se profundizaron temas que nutren en gran medida a los datathons organizados por la ASC-ARS, además de promover espacios interdisciplinarios con un enfoque centrado en la resolución de problemas. Para este año (2025) se plantea el tema de interfaces entre la estadística y la IA, dicho tema relaciona fuertemente a la estadística, la IA y el aprendizaje automático. (IASC-ARS Interim Conference 2024, s. f.)

Los datathons aportan, entre otras cosas, a la formación profesional, no solo del alumnado sino de los académicos, pues no hay restricción de participantes. Los estudiantes ponen a prueba el conocimiento adquirido apoyado de su creatividad, es por esto que la siguiente sección se presentan un análisis detallado las aportaciones que brindan a los estudiantes el poder participar en eventos de este tipo. Como inicio en la siguiente sección se centrará en explicar al lector los conceptos claves necesarios para el correcto entendimiento del tema, posterior a esto, se presentará un estado actual de ejemplos de

datathons y los principales productos obtenidos de estos y como es que las distintas ramas de la ciencia enriquecen estos eventos y como distintos sectores utilizan este tipo de eventos para realizar una caza de talentos y expandir la red de contactos.

2 Preliminares

Para el mejor entendimiento de lo que es un datathon, realizaremos una breve explicación de los conceptos clave de estos eventos, de acuerdo con Provost y Fawcett (2013), los datathons pueden ser vistos como eventos entretenidos que de alguna forma integran orgánicamente a la competencia con la colaboración.

Generalmente los equipos se forman con integrantes de distintas áreas del conocimiento, lo que lleva como resultado a un equipo altamente interdisciplinario, aunque usualmente formado en su mayoría por estudiantes de ciencias computacionales o de datos. El objetivo es, usualmente, con cantidades muy grandes de datos, resolver algún problema, proponer algún cambio, observar posibles conjeturas, etc. Debido a la multidisciplinariedad, se llegan a resultados que personas de un solo tipo de conocimiento no podrían llegar tan fácilmente. Un factor clave en la obtención del objetivo en la facilidad de los equipos a trabar en equipo, se puede tener a personas extremadamente brillantes, pero sino son capaces de expresar sus ideas y trabajar en equipo, probablemente no cumplan con la tarea asignada en el tiempo. La creatividad juega un papel fundamental en la obtención del producto final, pues en estos eventos el tiempo y como mencionamos la capacidad de trabajar en equipo, son el principal reto para vencer, eso sin contar a otros participantes.

Aportando al tema del aprendizaje basado en la experiencia, Kolb (2015) plantea que este aprendizaje se desarrolla de mejor manera cuando se colabora en un equipo en el cual exista una variedad creativa, o sea, se vincula de forma directa la experiencia con la práctica en equipo. Siguiendo esta idea, los datathons son eventos perfectos para el desarrollo de los estudiantes, pues poner a competir a personas con distinta formación y distintas competencias y capacidades, simula lo que en realidad ocurre en un ambiente laboral. De este modo los participantes ponen a prueba sus aptitudes y conocimiento para resolver un problema real y da como resultado la obtención de nuevo conocimiento, no solo individual sino que en el colectivo que es aplicado en un contexto real.

Ferrazzi (2014) observa a estos eventos como una oportunidad para realizar una red profesional o un “Networking” profesional, pues aquí se fomenta el construir y alguna forma mantener una nueva red de contactos en el ámbito académico. Además los datathons resultan ser una oportunidad de los estudiantes de ser vistos por caza talentos que usan a estos eventos como prueba para reclutar jóvenes que cumplan con los aspectos claves necesarios para ganar un datathon y de esta forma proporcionales oportunidades de desarrollo laboral.

Aunado a esto, Watanabe (2009) describe a la capacidad de resolver problemas como factor determinante que permite identificar a una persona que resalta entre otras. Dicha habilidad es clave en cualquier datathons, pues no siempre se trabaja con datos por decirlo

de alguna manera “limpios”, un paso clave en análisis de datos se centra en el acondicionamiento de los datos, pues no es siempre posible obtener datos organizados y es ahí en donde las capacidades técnicas, que obtuvo a lo largo de su formación, y la creatividad de una persona, propia de la persona, ayudan a la resolución del problema y por ende permiten resaltarla entre otras.

3 Aplicaciones de los Datathons en la Industria

Como se mencionó en la sección anterior, los datathons se han popularizado en las últimas décadas y cada vez más las industrias pertenecientes a distintas áreas ven a estos eventos como una oportunidad no solo de obtención de recursos humanos de gran valor, sino que también como una oportunidad de solucionar problemas relacionados a sus sector, mediante lo llamado “innovación abierta”, en donde básicamente, el sector privado, gobiernos, instituciones públicas y/o empresas colaboran con jóvenes talentos de distintas áreas del conocimiento, funcionando así como una inversión o un semillero de recursos humanos brillantes.

Como primer ejemplo, en el área de la salud el datathon presentado por Women in Data Science (WiDS), un evento organizado por y para mujeres involucradas en la ciencia de datos, se enfoca principalmente en la salud cerebral y emocional de las mujeres. Las participantes se centran en el análisis de imágenes, que son donadas a la ciencia y sin fines de lucro, de por ejemplo; resonancias magnéticas, tomografías o radiografías con el fin de proponer modelos y nuevas formas para el diagnóstico de enfermedades neurológicas, como el Parkinson. El evento es realizado con apoyo de Ann S. Bowers Women’s Brain Health Initiative quienes tienen el objetivo de mejorar la salud mental de las mujeres. A principios del año 2025, las participantes del Women in Data Science datathon desarrollaron modelos de machine learning para predecir diagnósticos de ADHD y el primer lugar logró predecir con base a patrones de conectividad cerebral, si el cerebro pertenecía a una mujer o a un hombre.

Ahora bien, en el mundo del dinero, en las finanzas, los datathons se consideran una herramienta muy útil para que las instituciones bancarias y de valores obtenga soluciones a problemas o respuestas a preguntas que no han podido responder. Los bancos usan estos eventos para mejorar su capacidad de detectar fallas en sus sistemas, identificación de posibles fraudes, mejorar en la atención al consumidor, optimizar inversiones y minimizar riesgos, por ejemplo en 2024 durante el Schonfeld Early Engagement Summit Ph.D. Datathon se obtuvieron modelos de machine learning que lograban predecir con gran precisión los retornos bursátiles de inversiones. Otro caso, quizá de los más famosos, en el ámbito de la economía es el datathon BBVA, un datathon que en algunas ediciones ha llevado a los estudiantes y profesionistas a trabajar durante días en la resolución de problemas, generalmente la institución presenta sets de datos reales, ya sean transacciones, patrones de uso, número de cuentas nuevas y cuales se cierran, horarios, atención al cliente, etc. En la edición llevada a cabo en el 2023, el equipo Mile Crew desarrolló un modelo en

machine learning que pudo predecir con certeza la fuga de clientes, esto permite a BBVA generar estrategias según el comportamiento histórico de cada cliente (BBVA, 2025).

Por otro lado, en el área de la energía y la lucha contra el cambio climático, está el Endesa Datathon, organizado por la empresa Endesa, una de las empresas energéticas más grandes de España. Este evento usa el poder del Big Data —esos conjuntos de datos enormes que a veces parecen imposibles de manejar— para atacar problemas del sector energético. En particular, se enfocan en cómo gestionar mejor la electricidad y hacer que las personas la usen de forma más eficiente. Supongamos que se requiere encontrar maneras para evitar el desperdicio de energía o bien, se pretende predecir cuánta electricidad va a necesitar una ciudad en un día caluroso, pues el análisis de datos empleando técnicas basadas en redes neuronales, vectores autoregresivos, etc. Resultan ser caminos viables.

De acuerdo con (La experiencia del Datathon, 2016), pagina web en donde se reportan los mejores momentos del datathon, este datathon muestra cómo los datos pueden cambiar por completo la industria energética, ayudándola a ser más sostenible y amigable con el planeta. Es como darle a la energía un cerebro digital para tomar decisiones más inteligentes. Entonces, ya sea en la salud, las finanzas, la energía o cualquier tema de interés social, los datathons no solo sirven para solucionar problemas del momento, sino que también crean puentes entre personas talentosas de afuera —como estudiantes o expertos independientes— y las industrias más importantes. Esto hace que trabajar juntos sea más dinámico y emocionante, y cambia la manera en que enfrentamos los desafíos de hoy en día, abriendo puertas a un futuro más innovador.

4 Desafíos y Recomendaciones para los Participantes

Participar en un datathon representa para los estudiantes de ingeniería en ciencia de datos, o a fines, una oportunidad única y valiosa de poner en práctica los conocimientos adquiridos en el aula, enfrentándose a problemas del mundo real que requieren soluciones innovadoras basadas en el análisis de interpretación de datos.

Los datathons permiten entre otras cosas, aplicar conceptos teóricos, fomenta el desarrollo de habilidades tanto técnicas como blandas, las cuales son vitales para el desarrollo profesional. Puede servir como guía para darse cuenta si se es capaz de trabajar en equipo y bajo presión constante y con la necesidad de comunicarse de forma efectiva, como se mencionó anteriormente, estos eventos sirven como semillero de talento, pero, sino se manejan adecuadamente podrían llegar a influir de forma negativa en el desarrollo del estudiante y en su experiencia del evento, es por ello que a lo largo de esta sección se presentan “tips” y recomendaciones a los estudiantes que pretenden ayudar a los estudiantes a que la experiencia no llegue a ser pesada.

Uno de los desafíos a los que los datathons se enfrentan es la estricta restricción de tiempo. A diferencia de proyectos académicos y/o laborales, en los que se puede extender durante semanas o meses, en un datathon los equipos deben conceptualizar, desarrollar e

implementar una solución completa en un plazo muy limitado, que suele oscilar entre unas pocas horas y un par de días a lo mucho. Esta presión temporal exige, entre otras cosas, una gestión eficiente del tiempo y una toma de decisiones rápidas, pero bien informadas, sustentadas con toda la data analizada. El autor Allen (2015) aborda este aspecto en su obra y sugiere varias estrategias clave para optimizar este proceso: distribuir las tareas entre los miembros del equipo desde el inicio del evento para evitar duplicidad de esfuerzos, establecer tareas intermedias que permitan evaluar el progreso y así ajustar el rumbo si fuese necesario, de esta forma evitar caer en la trampa de perfeccionismo dedicando un tiempo excesivo a un solo componente del problema, como el diseño de un modelo o la limpieza de datos. Esta planificación inicial y el seguimiento constante son fundamentales para garantizar que el equipo avance de manera coordinada hacia una solución viable dentro del plazo establecido.

En los datathons, los participantes forman equipos multidisciplinarios, es decir, grupos donde cada integrante puede o no tener una formación distinta, como programación, estadística o diseño y no siempre del área de las ciencias exactas. Sin embargo, trabajar en equipo no siempre es sencillo. Si no se definen bien los roles de cada persona ni se establecen formas claras de comunicación, pueden surgir problemas como malentendidos, conflictos e inclusive una desorganización. Imaginemos un equipo de fútbol donde nadie sabe quién es el delantero, quien es el defensa, la posición del porte no está cubierta, etc: sería un caos. Para evitar esto, Lencioni (2002) sugiere algunas ideas prácticas: elegir un líder de equipo que organice y dirija al grupo, fomentar que todos hablen abiertamente y participen en las decisiones, y dividir las tareas según las fortalezas de cada integrante. Por ejemplo, si alguien es bueno analizando números, podría encargarse de los cálculos, mientras que otro que sabe comunicar ideas podría presentar los resultados.

Por otro lado, el análisis de datos, el corazón de un datathon puede llegar a ser todo un desafío. Los participantes suelen enfrentarse a conjuntos de datos masivos y complicados, como listas enormes de números, fechas o información desordenada. Si no se tiene experiencia en preparar esos datos —lo que incluye limpiarlos, explorarlos y transformarlos—, siguiendo metodologías típicas ETL (extracción, transformación y carga, por ejemplo), el trabajo se vuelve lento y los resultados pueden no ser buenos. Es como intentar cocinar sin lavar los ingredientes primero: el plato no va a salir bien. Para hacerlo más fácil, Mayer-Schönberger y Cukier (2013) explican que lo primero es entender rápidamente de qué trata el conjunto de datos (dataset), es decir, qué información tiene y cómo está organizada.

Una vez comprendido el dataset, hay que usar herramientas amigables y eficientes para analizarlos, en Python existen herramientas como Pandas que funciona como una especie de navaja suiza para organizar y manipular datos; para trabajar con bases de datos enormes, el conocimiento en SQL o el uso de gestores de bases de datos (DBMS) es fundamental. También, es clave definir estrategias de preprocesamiento para manejar problemas comunes, como datos faltantes (por ejemplo, si en una tabla no aparece la edad de una persona) o inconsistencias (como una fecha imposible o valores fuera del rango de valores “normales”). Con estas tácticas, el equipo puede avanzar más rápido y obtener resultados confiables que realmente resuelvan el problema del datathon.

5 Conclusiones

Es así como llegamos a la quinta y última sección en la que podemos resumir el capítulo de la siguiente forma: Los datathons son una categoría de los hackathons que llevan existiendo desde hace más de dos décadas, pero es desde que la tecnología maduró cuando llegó el boom de estos. Los datathons son una mezcla de competición y colaboración muy particular, existen diversos tipos de datathons cada uno con un tema muy particular y problema a resolver. Ya sea del ámbito de la salud, seguridad, energía, economía, etc.

Los datathons reúnen a equipos conformados por estudiantes y/o profesionales de distintas áreas, lo que permite la colaboración interdisciplinaria, y el trabajo en equipo. La creatividad, adaptabilidad y la capacidad de trabajar en equipo es fundamental para tener una experiencia agradable durante el evento. Debido al tema de la multidisciplinariedad, muchas instituciones ven a los datathons como semillero de talento emergente y en estos eventos es muy común ver a cazatalentos u ofrecer premios tanto en efectivo como pasantías o estancias en dichos institutos.

Es por eso por lo que después de profundizar en el tema de que es y en qué consiste un datathon, se presenta una sección en la que se realizan recomendaciones a los estudiantes interesados en participar en estos eventos. Esto con el fin de que logren desarrollarse profesionalmente y aumenten su capacidad de trabajar colaborativamente. De esta forma podemos ver a los datathons no solo como una competencia, sino como una oportunidad de establecer conexiones con redes de colaboradores y poder prepararse mejor para los retos del mundo real laboral.

Referencias

- Allen, D. (2015). *Getting things done: The art of stress-free productivity* (Revised edition). Penguin Books.
- Anslow, C., Brosz, J., Maurer, F., & Boyes, M. (2016). Datathons: An Experience Report of Data Hackathons for Data Science Education. *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, 615-620. <https://doi.org/10.1145/2839509.2844568>
- BBVA. (2025, febrero 22). *BBVA Perú*. <https://www.bbva.pe/personas/data-challenge.html>
- Data Analysis Competition | IASC-ISI*. (s. f.). Recuperado 8 de febrero de 2025, de <https://iasc-isi.org/data-analysis-competition/>
- Ferrazzi, K. (with Raz, T.). (2014). *Never Eat Alone, Expanded and Updated: And Other Secrets to Success, One Relationship at a Time*. Crown/Archetype.
- García, J., Molina, J., Berlanga, A., Patricio, M., Bustamante, A., & Padilla, W. (2018). *Ciencia de Datos: Técnicas Analíticas y Aprendizaje Estadístico*. Alfaomega.
- Home*. (s. f.). DSSV - ECDA 2023. Recuperado 16 de febrero de 2025, de <https://iasc-isi.org/dssv-ecda2023/>
- IASC-ARS Interim Conference 2024*. (s. f.). IASC-ARS Interim Conference 2024. Recuperado 16 de febrero de 2025, de https://webpageprodvm.ntu.edu.tw/IASC-ARS_Interim_2024/Default.aspx
- Kolb, D. A. (2015). *Experiential learning: Experience as the source of learning and development* (Second edition). Pearson Education, Inc.
- La experiencia del Datathon*. (2016, mayo 15). Endesa. <https://www.endesa.com/es/la-cara-e/talento/experiencia-datathon>
- Lemus-Delgado, D., & Navarro, R. P. (2020). Ciencia de datos y estudios globales: Aportaciones y desafíos metodológicos. *Colombia Internacional*, 102, Article 102. <https://doi.org/10.7440/colombiaint102.2020.03>
- Lencioni, P. (2002). *The five dysfunctions of a team: A leadership fable* (1st ed). Jossey-Bass. .
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking* (First edition). O'Reilly.
- Watanabe, K. (2009). *Problem solving 101: A simple book for smart people*. Portfolio.
- WiDS Datathon 2025*. (s. f.). Recuperado 23 de febrero de 2025, de <https://kaggle.com/widsdatathon2025>

Capítulo 6

El impacto de la desdemocratización de la ciencia de datos en el desarrollo de la educación y la ciencia

Berenice Rodríguez-Pedroza, Mauro Alberto López-Muñoz, María Teresa Torrijos-Muñoz

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, Puebla, México
berenice.rodriguezp@correo.buap.mx,
mauroa.lopez@correo.buap.mx,
teresa.torrijos@correo.buap.mx

Resumen. En los últimos años, México ha empezado a incluir la ciencia de datos dentro de sus programas educativos debido a la creciente necesidad de analizar grandes volúmenes de datos que contribuyan a la toma de decisiones informadas en el sector privado y público. En contraparte, la desdemocratización de la información limita el acceso a recursos y herramientas tecnológicas que impactan en el desarrollo de la educación de dicha disciplina. Por lo cual el presente artículo, busca realizar una revisión exploratoria sobre las circunstancias actuales que conllevan a una inequidad de los recursos para la enseñanza en ciencia de datos; identificando como factores principales: la carencia de infraestructura tecnológica, la falta de marcos institucionales, la desigualdad en el acceso a los datos e inclusive, la diferencia idiomática en los contenidos educativos que abordan los últimos avances en el campo de la ingeniería.

Palabras Clave: Ciencia de datos, brecha digital, acceso a los datos.

1 Introducción

La Ciencia de Datos permite extraer información significativa a través de la identificación de patrones y tendencias de un problema en específico con la finalidad de coadyuvar a la toma de decisiones sea desde un entorno empresarial o académico.

Actualmente diversas universidades, tanto privadas como públicas, han comenzado a capacitar profesionistas en ciencias de datos que ayuden a la mejora de las operaciones en diferentes sectores tales como el marketing, finanzas, salud, manufactura, logística, entre otros. Para llevar a cabo el análisis de datos se requiere un enfoque multidisciplinario abordando áreas del conocimiento tales como la matemáticas, inteligencia artificial, estadística y ciencias computacionales (Crespo, 2022).

A pesar del interés por incluir un enfoque multidisciplinario, la brecha digital sigue siendo un obstáculo para el aprendizaje con tecnología dentro de América Latina (Rodríguez, 2021). Además, el poder adquisitivo influye significativamente en el

acceso a la digitalización con fines académicos. Se estima que el 18 % de los adolescentes latinoamericanos de 15 años presenta desventajas socioeconómicas que les impiden el acceso al internet, en contraste con un promedio inferior al 2 % en los países de la OCDE. Además, menos del 10 % de las personas en Chile, Ecuador, México y Perú tienen un nivel completo de alfabetización, cálculo y habilidades en resolución de problemas en entornos con alta presencia tecnológica (Minea, 2020).

De acuerdo con el último Monitor Global de Educación de Ipsos, se obtuvo que el 65 % de los encuestados de 29 países consideran importante que los docentes se capaciten en inteligencia artificial. En particular, México registra una cifra de un 79 % de interés en incluir dicha disciplina en las aulas (Martín, 2023). En contraparte, la inequidad de la información limita el acceso a recursos y herramientas tecnológicas que impactan en el desarrollo de la educación. Por tal motivo, este estudio tiene como objetivo responder a la pregunta central de investigación: ¿cuáles son los factores principales que actualmente conllevan a desdemocratización de la ciencia y la educación?

El presente trabajo comienza con una sección dedicada a la metodología, la cual describe los criterios de inclusión y el enfoque PRISMA para la selección de los artículos. Como segunda sección se presenta la revisión exploratoria que aborda los factores principales relacionados con la inequidad en el acceso a la información digital, y su impacto en la educación y ciencia. Finalmente se abordan las conclusiones del estudio.

2 Metodología

En este artículo se presenta una revisión exploratoria cuyo objetivo es ofrecer un panorama inicial sobre la desdemocratización de la ciencia de datos en el desarrollo de la educación y la ciencia. Por lo cual, para la realización del presente trabajo se siguieron las fases propuestas por Arkey y O'Malley (Fernández, 2020) descritas en la tabla 1.

Tabla 1. Fases de la Revisión Sistemática Exploratoria por Arkey y O'Malley

Fases	Descripción
Fase 1	Elaboración de la pregunta de investigación.
Fase 2	Establecimiento de los criterios de inclusión y exclusión y búsqueda sistemática.
Fase 3	Revisión y selección de artículos
Fase 4	Extracción de datos
Fase 5	Análisis y reporte de resultados.

De esta manera, se comenzó con la consulta en buscadores de literatura académica, en particular Google Académico, Redalyc, Scielo y ScienceDirect.

Los criterios de inclusión de la presente revisión exploratorio se basaron en textos en español o inglés entre los años 2020 a 2025, los cuales abordaran los desafíos de la ciencia de datos que han tenido un impacto directo tanto en la educación como en el desarrollo científico en los últimos cinco años y en particular, en América Latina. Para ello, se definieron búsquedas booleanas, tal como se observa en la tabla 2.

Tabla 2. Búsqueda aplicada y sus respectivos resultados

Bases de datos	Búsqueda booleana	Resultados de búsqueda	Artículos identificados	Artículos seleccionados
Google Académico	"ciencia de datos" and "educación"	40	8	2
	"Brecha digital" or "industria 4.0" and "ciencia de datos"	40	6	3
Redalyc	"ciencia de datos" and "educación"	40	3	1
	"Brecha digital" or "industria 4.0" and "ciencia de datos"	40	4	1
Scielo	"ciencia de datos" and "educación"	15	2	1
	"Brecha digital" or "industria 4.0" and "ciencia de datos"	2	0	0
Science Direct	"ciencia de datos" and "educación"	50	4	3
	"Brecha digital" or "industria 4.0" and "ciencia de datos"	50	3	3
Total			30	14

Al aplicar la búsqueda booleana, se consideró como máximo los primeros cincuenta resultados, debido a que se observó que los artículos posteriores se alejaban del foco central de la presente investigación. De los resultados obtenidos, se corroboró si los títulos se ajustaban a los temas centrales establecidos. Tal como se describe en la tabla 1, se identificaron 30 artículos candidatos. Posteriormente, en la figura 1, se muestra el enfoque PRISMA que describe el flujo del proceso de selección de los artículos.

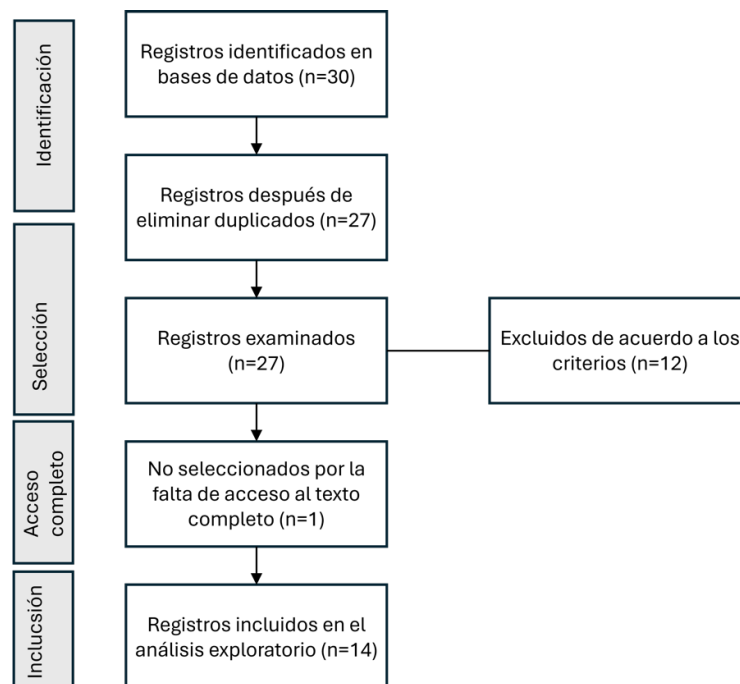


Figura 1. Diagrama de flujo PRISMA para la revisión sistemática de la literatura

Mediante el enfoque PRISMA se procedió a identificar la duplicidad de artículos y posteriormente, se realizó una exploración más detallada del contenido de cada artículo, seleccionando los documentos que abordaban los desafíos de la descentralización de la ciencia de datos, así como su impacto en el desarrollo académico y científico.

En la tabla 3 se muestra una tabla que sintetiza la temática central de cada artículo seleccionado.

Tabla 3. Artículos seleccionados

Artículo	Título	Temática central
Acosta, M. E. (2024)	<i>La Industria 4.0: Transformando el futuro de la Construcción</i>	La brecha digital y el acceso desigual a la tecnología como desafío dentro de la Industria 4.0
Barrientos, O. T., Hernández, M. R., & Alva, A. D. (2024)	<i>Importancia de la Inteligencia Artificial y la Ciencia de Datos en el sector Educativo del Siglo XXI</i>	El impacto de la ciencia de datos y la inteligencia artificial en la educación

Cinnamon, J. (2020).	<i>Data inequalities and why they matter for development. Information Technology for Development</i>	Desigualdad en el acceso a la información digital en países en desarrollo.
Ciupercă, E. M., Donnelly, N., Gartland, A., & Stanciu, A. (2022).	<i>The Digital Divide in Education-Macrocultural Comparative Analysis between Ireland and Romania</i>	Estudio sobre el impacto de la pandemia en el acceso a las herramientas digitales.
Connolly, G., Costa-Font, J., & Srivastava, D. (2025).	<i>Did COVID-19 reduce the digital divide? A systematic review</i>	Análisis sobre el acceso a la infraestructura digital debido a la pandemia del COVID-19
Contreras-Hernández, A., Equihua, M., Pérez-Maqueo, O., Equihua, J., & Morandin-Ahuerma, I. (2023).	<i>Vinculación institucional, educación y tecnología. Claves para atender la crisis socio-ecosistémica.</i>	La accesibilidad de los datos para el beneficio de una sociedad sustentable.
Fadillah, M. A., & Siregar, F. A. (2025).	<i>Bibliometric mapping of data science in education: Trends, benefits, challenges, and future directions.</i>	Revisión sistemática sobre las necesidades, retos y limitaciones en la enseñanza de la ciencia de datos
Fernandez, C., Freitas, J. A., Blikstein, P., & de Deus Lopes, R. (2025).	<i>The design space of visualization tools for data science education: Literature review and framework for future designs</i>	Revisión exploratoria sobre funciones de visualización de datos para brindar acceso a entornos educativos en ciencia de datos.
García-Vargas, G. R., Parés-Hipólito, J., Serrano, M., & Martínez, L. A. R. (2024).	<i>El uso de la Inteligencia Artificial y la Ciencia de Datos en la Educación: Preocupaciones éticas</i>	Las cuestiones éticas que derivan del uso de la ciencia de datos
Lemus-Delgado, D., & Pérez Navarro, R. (2020)	<i>Ciencia de datos y estudios globales: aportaciones y desafíos metodológico</i>	Las implicaciones que presenta la ciencia de datos en estudios globales
Martín, F. S., Valdivia, G. N. V., & Castro, R. R. (2023)	<i>Panorama actual de la ciencia de datos en la educación superior en México</i>	Incorporación de la ciencia de datos en las universidades mexicanas desde el 2018 hasta el 2023
Masood, K. M. (2022).	<i>English Competencies and Challenges for Data Science and Cyber Security Students at Al Istiqlal University</i>	La necesidad del idioma inglés en la formación de ingenieros en ciencias de datos.
Moletsane, R. (2025).	<i>The Impact of Digital Divide on Women: A Rural Community Case</i>	El impacto de la brecha digital en el desarrollo profesional de

		las mujeres dentro de las comunidades rurales.
Okoye, K., Campos, E., Das, A., Chakraborty, V., Ghosh, M., Chakrabarti, A., & Hosseini, S. (2025).	<i>Impact of digitalized-education upon sustainable education and practice: A systematic review and meta-analysis of literature based on pre-intra-and-post pandemic and rural education development</i>	Análisis de las implicaciones del uso de tecnologías digitales en el quehacer pedagógico, en particular en las zonas rurales.

3 Revisión exploratoria

En esta sección se exploran los principales factores que, en los últimos cinco años, han llevado a una inequidad de los recursos en la enseñanza dirigida a ciencia de datos. Para ello, se presenta una discusión sobre las ideas principales de los artículos seleccionados.

3.1 Acceso a la infraestructura tecnológica

En primer lugar, Martín et al. (2023) examinan el desarrollo de la ciencia de datos en los últimos años, en particular en el territorio mexicano. Esta disciplina ha tenido gran relevancia debido a sus aplicaciones en la toma de decisiones de diversos sectores tales como las finanzas, la salud, el servicio al cliente, la industria manufacturera, la investigación científica y la educación. Por lo cual, los autores destacan la necesidad de fortalecer colaboraciones entre instituciones educativas y empresas para el intercambio de conocimiento.

En relación con el ámbito empresarial, Acosta (2024) destaca que para favorecer los avances de la Industria 4.0, en la cual intervienen áreas como la ciencia de datos y la inteligencia artificial, se requiere atender la brecha digital entre regiones. Por su parte Fadillah (2025) complementa este análisis al destacar que la capacidad económica influye significativamente en la adquisición del hardware y software destinado a aplicaciones de Big Data, además de que una solidez financiera permite contribuir a mantenimientos y actualizaciones periódicas en la infraestructura. Por lo cual, la brecha digital puede generar desigualdades entre países desarrollados y en vías de desarrollo tecnológico, limitando así las oportunidades crecimiento industrial y científico.

De manera similar, Barrientos et al. (2024) apoyan la importancia de reducir la brecha digital con el objetivo de que el uso de las tecnologías sea más inclusivo en las instituciones. A su vez, el autor también explora el impacto de la investigación en ciencia de datos en el ámbito educativo; identificando que países como México y Ecuador a pesar de contar con publicaciones en el área de ciencia de datos, aún presentan desafíos tanto en la infraestructura de investigación como en el apoyo institucional a diferencia de países como España. El autor Fernández (2025) comparte la misma visión, al destacar la importancia de orientar el acceso a herramientas tecnológicas en la enseñanza de Ciencia de Datos.

Fuera del contexto urbano, autores como Moletsane y Okoye resaltan el impacto la brecha digital en la enseñanza de las comunidades rurales quienes presentan mayor dificultad de un acceso internet estable tanto en sus hogares como en sus escuelas, en particular dicha problemática rural afecta en mayor proporción a la población femenina.

Como último punto y no menos importante, Connolly y Ciupercă hacen énfasis de cómo la tecnología tuvo un crecimiento acelerado durante la pandemia, debido a la necesidad de adaptar las funciones empresariales, educativas en un contexto de aislamiento. A su vez, el acceso a las herramientas digitales permitió un intercambio de información crucial para la toma de decisiones sanitarias. Por ello, es necesario evaluar modelos de financiamiento como subsidios y programas de ayuda gubernamental para la adopción de nuevas tecnologías emergentes.

3.2 Acceso a los datos, modelos y literatura especializada

García-Vargas et al. (2024) también comparten la necesidad de un acceso equitativo a las nuevas tecnologías. Además, incluyen a los algoritmos sesgados, el determinismo tecnológico y la erosión humana como parte de las preocupaciones éticas relacionadas con la ciencia de datos. El determinístico tecnológico y la erosión de un agente humano se refieren a la falta de transparencia en los procesos algoritmos en sistemas cerrados, lo que puede generar decisiones automatizadas sin la supervisión humana en situaciones complejas. Por su parte, el sesgo algorítmico implica un entrenamiento inadecuado de los datos, lo cual podría conllevar a una representación parcial de una sociedad, afectando a grupos minoritarios o vulnerables. Este último punto también ha sido abordado por Lemus-Delgado et al. (2020), quienes enfatizan que, al existir una brecha entre el diseño de los algoritmos y las implicaciones éticas en su respectivo funcionamiento podrían afectar, incluso, a sociedades enteras.

Adicionalmente, la carencia de marcos regulatorios enfocados en la transparencia puede conllevar a un sesgo en la producción de datos y esto a su vez, a una representación desigual de la sociedad como lo menciona Cinnamon (2020). El autor también expone una desigualdad educativa debido a un acceso inequitativo de los datos. La posible incapacidad de acceder a los datos puede radicar en factores tales como la pobreza, la falta de infraestructura tecnológica y barreras idiomáticas. En concordancia con este último punto, Masood (2022) destaca que el idioma inglés influye significativamente en el rendimiento académico en carreras como la ciencia de datos y la ciberseguridad, por lo cual la falta de conocimiento en dicho idioma puede generar dificultades para acceder a literatura especializada o técnica.

Por otra parte, Contreras-Herández et al. (2023) abordan la importancia de los datos en el marco del Open Data Chapter, el cual establece que los datos deben ser abiertos por defecto, oportunos y exhaustivos, accesibles y utilizables, comparables e interoperables, orientados a mejorar la gobernanza y la participación ciudadana, así como el desarrollo incluyente y la innovación. De manera que la ciencia de datos, al procesar grandes volúmenes de información, coadyuve a la toma de decisiones y la generación de nuevo conocimiento basados en la transparencia y la accesibilidad.

3.3 Recomendaciones

Con base en el presente estudio se refleja la necesidad de fomentar apoyos gubernamentales que contribuyan al crecimiento de infraestructura tecnológica con conectividad estable en universidades tanto en zonas rurales como zonas urbanas.

Al mismo tiempo para coadyuvar a un acceso equitativo de los modelos actuales en ciencia de datos y a la literatura especializada es fundamental contribuir a la publicación en revistas y repositorios de acceso abiertos. A su vez dichos materiales digitales cuenten con opciones de traducción en el idioma de la región y de ser necesario, considerar un trabajo colaborativo con traductores de lenguas indígenas para la difusión de las disciplinas básicas que se relacionan con la enseñanza de la ciencia de datos, fomentando así una participación más activa en el sector estudiantil tanto en áreas urbanas como rurales.

Adicionalmente, la colaboración entre centros de investigación, universidades e industria permitiría en gran medida que las instituciones educativas accedan a casos de estudios de problemáticas actuales, fortaleciendo así el pensamiento crítico-analítico de sus estudiantes y al mismo tiempo, las empresas reciban asesoramiento en modelos y técnicas vanguardistas publicadas en revistas científicas dirigidas a la ciencia de datos.

Finalmente, dentro de este proceso de colaboración también es fundamental que existan marcos regulatorios enfocados a la supervisión en la toma de decisiones de los modelos y la transparencia en la respuesta de estos. Al mismo tiempo, dichos marcos deben ser supervisados y cumplidos de manera conjunta entre organismos federales, instituciones y el sector privado.

4 Conclusiones

Durante la revisión exploratoria se obtuvieron más de 100 resultados de búsqueda, la mayoría de los artículos se centraban en identificar herramientas de inteligencia artificial para la enseñanza y en cómo la ciencia de datos identifica tendencias en la educación o la industria. Por consiguiente, sólo fueron seleccionado catorce artículos relevantes para el presente estudio, lo cual indica que el impacto de la desdemocratización de la ciencia de datos en el contexto académico sigue siendo un tema poco explorado.

Por lo tanto, existe un área de oportunidad en la planeación de estrategias que permitan ampliar el acceso abierto a la información de vanguardia bajo la definición de marcos institucionales y al mismo tiempo favorecer una educación con equidad en herramientas tecnológicas; lo cual tendría un impacto fundamental en la generación de nuevo conocimiento tanto en la enseñanza en ciencias de datos como en su desarrollo científico.

Referencias

- Acosta, M. E. (2024). *La Industria 4.0: Transformando el futuro de la Construcción*. Robótica: promesas, utopía o futuro distópico, 2.
- Barrientos, O. T., Hernández, M. R., & Alva, A. D. (2024). *Importancia de la Inteligencia Artificial y la Ciencia de Datos en el sector Educativo del Siglo XXI*. Transforming Education, 89.
- Cinnamon, J. (2020). *Data inequalities and why they matter for development*. *Information Technology for Development*, 26(2), 214-233.
- Ciupercă, E. M., Donnelly, N., Gartland, A., & Stanciu, A. (2022). The Digital Divide in Education-Macro cultural Comparative Analysis between Ireland and Romania. *IFAC-PapersOnLine*, 55(39), 99-104
- Connolly, G., Costa-Font, J., & Srivastava, D. (2025). Did COVID-19 reduce the digital divide? A systematic review. *Health Policy and Technology*, 100979.
- Contreras-Hernández, A., Equihua, M., Pérez-Maqueo, O., Equihua, J., & Morandin-Ahuerma, I. (2023). *Vinculación institucional, educación y tecnología. Claves para atender la crisis socio-ecosistémica*. *Perfiles educativos*, 45(180), 140-157.
- Crespo, F., Alves, T., & Soto, M. (2022). *Ciencia de Datos, Inteligencia Artificial, y sus impactos sobre la sociedad*. *Observatorio Económico*, (169), 9-11.
- Fadillah, M. A., & Siregar, F. A. (2025). Bibliometric mapping of data science in education: Trends, benefits, challenges, and future directions. *Social Sciences & Humanities Open*, 11, 101600
- Fernandez, C., Freitas, J. A., Blikstein, P., & de Deus Lopes, R. (2025). The design space of visualization tools for data science education: Literature review and framework for future designs. *International Journal of Child-Computer Interaction*, 43, 100698.
- Fernández-Sánchez, H., King, K., & Enríquez-Hernandez, C. B. (2020). *Revisiones Sistemáticas Exploratorias como metodología para la síntesis del conocimiento científico*. *Enfermería universitaria*, 17(1), 87-94.
- García-Vargas, G. R., Parés-Hipólito, J., Serrano, M., & Martínez, L. A. R. (2024). *El uso de la Inteligencia Artificial y la Ciencia de Datos en la Educación: Preocupaciones éticas*. Transforming Education, 127.
- Lemus-Delgado, D., & Pérez Navarro, R. (2020). *Ciencia de datos y estudios globales: aportaciones y desafíos metodológicos*. *Colombia Internacional*, (102), 41-62.

- Martín, F. S., Valdivia, G. N. V., & Castro, R. R. (2023). *Panorama actual de la ciencia de datos en la educación superior en México*. South Florida Journal of Development, 4(4), 1705-1714.
- Masood, K. M. (2022). *English Competencies and Challenges for Data Science and Cyber Security Students at Al Istiqlal University*. The Creative launcher, 7(5), 48-68.
- Minea, A. (Ed.). (2020). *Making the Most of Technology for Learning and Training in Latin America*. OECD Publishing.
- Moletsane, R. (2025). The Impact of Digital Divide on Women: A Rural Community Case. *Procedia Computer Science*, 263, 791-797.
- Okoye, K., Campos, E., Das, A., Chakraborty, V., Ghosh, M., Chakrabarti, A., & Hosseini, S. (2025). Impact of digitalized-education upon sustainable education and practice: A systematic review and meta-analysis of literature based on pre-intra-and-post pandemic and rural education development. *Sustainable Futures*, 100851.
- Rodríguez-Alegre, L. R., Trujillo-Valdiviezo, G., Egusquiza-Rodríguez, M. J., & López-Padilla, R. D. P. (2021). Revolución industrial 4.0: La brecha digital en Latinoamérica. *Revista arbitrada interdisciplinaria Koinonia*, 6(11), 147-162.

Capítulo 7

El rol del Ingeniero en Ciencia de Datos en un entorno de creciente democratización de la disciplina

Carlos Leopoldo Carreón Díaz de León, María Teresa Torrijos Muñoz, Mireya Tovar Vidal y Juan Carlos Conde Ramírez

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, México

`carlos.carreon@correo.buap.mx`, `teresa.torrijos@correo.buap.mx`,
`mireya.tovar@correo.buap.mx`, `Carlos.conder@correo.buap.mx`

Resumen. Hoy en día la ciencia de datos vive un proceso de apertura muy evidente, impulsado por el crecimiento de plataformas no-code y de herramientas que automatizan tareas. Esto ha cambiado, de forma paulatina pero clara, el papel que antes tenía el ingeniero en este campo. El análisis que se presenta busca mostrar cómo esta democratización no solo influye en la práctica profesional, sino también en la manera en que se toman decisiones apoyadas en información. Durante el desarrollo del tema aparecen tanto desafíos como oportunidades: asegurar la calidad de los datos y de los modelos, construir soluciones que puedan escalar, fomentar un uso responsable y, además, lograr que dichos modelos se integren de manera efectiva en las estructuras de una organización. Otro punto central es la dimensión ética: cómo manejar la información y reducir posibles sesgos en los algoritmos.

Palabras Clave: Ciencia de datos, Democratización, Calidad de datos.

1 Introducción

En los últimos diez años la ciencia de datos explotó en popularidad. De ser un tema que solo dominaban unos cuantos especialistas, pasó a convertirse en algo cotidiano en áreas como las finanzas, el marketing, la salud, la industria e incluso el gobierno (Jordan & Mitchell, 2015). Y no es casualidad: cada vez enfrentamos más información y urge encontrar patrones y tendencias dentro de esos volúmenes gigantescos de datos. Para lograrlo han aparecido herramientas mucho más versátiles, capaces de recopilar, almacenar y mostrar la información de formas que antes parecían impensables (Zhang, Yang, Chen & Li, 2018).

Pero este crecimiento no solo se trata de tecnología. También ha cambiado quién puede usarla. Lo que antes estaba en manos de expertos ahora empieza a llegar a muchas más personas, un proceso que suele llamarse “democratización” (Cao, 2017).

Las plataformas no-code y low-code son el mejor ejemplo. Con ellas, alguien sin grandes conocimientos técnicos puede aplicar métodos de análisis o incluso correr un modelo de aprendizaje

automático (González & Chen, 2021). Esto abre la puerta a beneficios muy claros: más innovación, más gente resolviendo problemas con datos. Claro que no todo es positivo. También surgen dudas sobre la calidad de esas soluciones, si los modelos realmente pueden crecer y si los principios de ingeniería no se están dejando de lado (Sculley et al., 2015).

Por eso, lejos de volverse innecesario, el ingeniero en ciencia de datos tiene hoy un papel más estratégico que nunca. Es la persona que diseña arquitecturas confiables, que cuida que los procesos funcionen bien y que guía a quienes se suman a este ecosistema (Amershi et al., 2019). En este sentido, la expansión de la disciplina no borra su rol, sino que lo transforma. El reto está en asumir estas nuevas responsabilidades, aprovechar las oportunidades y, sobre todo, no perder de vista la ética y la lucha contra los sesgos en los algoritmos.

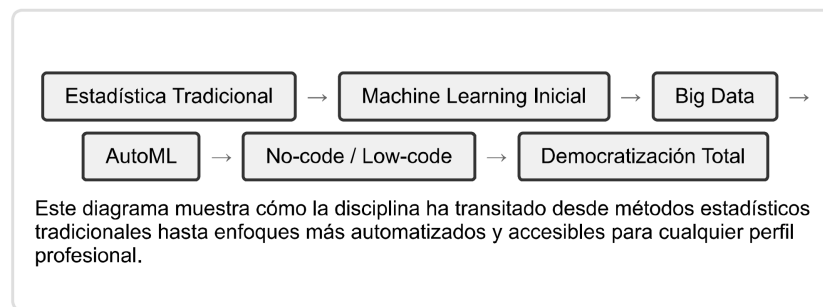


Figura 1 Flujo de la Evolución de la Ciencia de Datos.

2. Antecedentes de la democratización de la ciencia de datos

Hoy en día hablamos de ciencia de datos como un campo que se nutre de varias raíces: la estadística, el aprendizaje automático y la computación de alto rendimiento. Hacia mediados de la década de 2010 empezó a crecer la inquietud entre investigadores: ¿por qué estas herramientas deberían quedar solo en manos de unos cuantos expertos? (Dhar, 2013; Kandel, Heer, Plaisant & Kennedy, 2018). Ese cuestionamiento abrió el camino a una idea que poco a poco ganó terreno: la “democratización”. En pocas palabras, hacer que el acceso sea más amplio, derribar barreras y permitir que la innovación venga también de públicos diversos (Cao, 2017).

En este camino hay varios hitos que marcan la diferencia. Uno de los más visibles ha sido la llegada de las herramientas no-code y low-code. Con ellas, personas sin conocimientos técnicos avanzados pueden construir modelos predictivos, manipular datos o crear visualizaciones interactivas (González & Chen, 2021). La lógica es sencilla: interfaces gráficas que funcionan con plantillas y procesos de arrastrar y soltar, lo que hace que gran parte del trabajo se vuelva más intuitivo y automático.

Otro paso decisivo ha sido la popularización del AutoML (Automated Machine Learning). Esta tendencia cambió la manera de elegir modelos, ajustar parámetros y evaluar algoritmos (He, Zhao & Chu, 2021). Claro que, aunque facilita mucho el trabajo y ahorra recursos, también tiene su lado

delicado: fomenta la dependencia de sistemas que parecen “cajas negras” y que a veces ocultan los fundamentos matemáticos y computacionales que sostienen sus resultados.

Un aspecto que cambió el panorama fue el acceso a infraestructura digital más barata y disponible. Con la nube creciendo y el precio del cómputo y el almacenamiento bajando, la ciencia de datos dejó de ser exclusiva de grandes corporaciones. De pronto, también podían sumarse pequeñas empresas e incluso personas trabajando por su cuenta (McKinney, 2017). Ese escenario abrió muchas posibilidades, pero no todo son ventajas. También aparecieron dudas sobre si las soluciones realmente son confiables, sobre la forma en que se usan y sobre qué tan lejos se pueden llevar sin perder calidad (Sculley et al., 2015).

3. La democratización y su impacto en la práctica profesional

Disciplinas como la estadística, el aprendizaje automático y la computación de alto rendimiento dieron forma a lo que hoy llamamos ciencia de datos. A partir de la segunda mitad de la década de 2010 comenzó a discutirse la necesidad de abrir estas técnicas a más gente y no solo a expertos en el área (Dhar, 2013; Kandel, Heer, Plaisant & Kennedy, 2018). En medio de ese debate apareció con fuerza el término “democratización”, usado para describir la idea de quitar trabas de entrada y estimular nuevas formas de innovación (Cao, 2017).

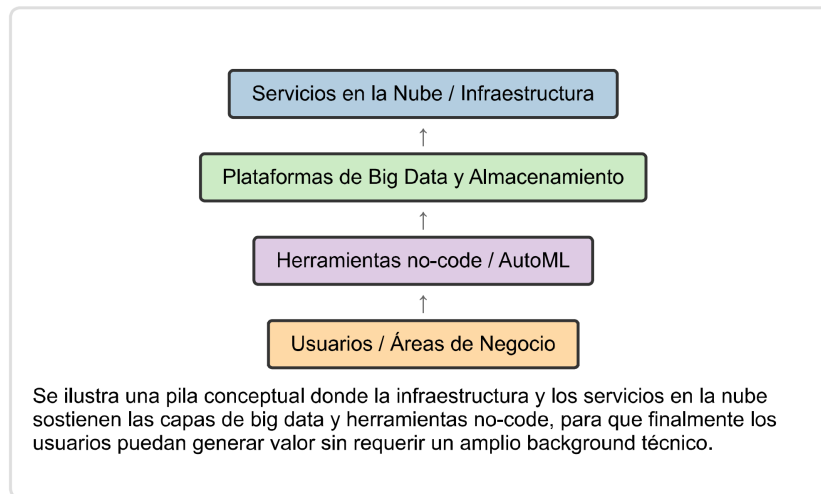


Figura 2 Arquitectura conceptual de la democratización.

La expansión de la ciencia de datos no se explica por un solo motivo, sino por varios cambios que se fueron acumulando. Uno de ellos fue la aparición de herramientas no-code y low-code. Estas plataformas, con menús visuales y funciones de arrastrar y soltar, acercaron el análisis a personas que

nunca habían programado y aun así pudieron crear modelos o visualizar datos (González & Chen, 2021).

Otro cambio vino con el AutoML. La automatización de tareas como elegir modelos o ajustar parámetros redujo tiempos de trabajo de manera notable (He, Zhao & Chu, 2021). El beneficio fue evidente, aunque también surgió un riesgo: confiar demasiado en sistemas opacos y perder de vista los fundamentos que hacen que funcionen.

El tercer giro estuvo en la infraestructura. La nube creció, el almacenamiento se abarató y el cómputo dejó de ser un recurso exclusivo de grandes corporaciones. Eso abrió la puerta a que pequeñas empresas e incluso personas solas pudieran emprender proyectos antes impensables (McKinney, 2017). Con esa apertura llegaron más aplicaciones posibles, pero también discusiones sobre la calidad de lo que se produce, la forma en que se usa y los límites para escalarlo de manera responsable (Sculley et al., 2015).

4. El papel del ingeniero en ciencia de datos en un entorno democratizado

La expansión de herramientas accesibles no elimina la figura del ingeniero en ciencia de datos. Al contrario, redefine su papel y lo coloca en un punto estratégico. Más allá de programar modelos, su aporte está en guiar a los equipos, dar solidez a las soluciones y asegurar que lo que se construye realmente genera valor en la organización.

4.1 Diseño de soluciones y escalabilidad

Crear un prototipo con una plataforma no-code es sencillo y rápido, pero convertirlo en un sistema estable que funcione en producción es otra historia. Ahí entra el ingeniero, que debe pensar arquitecturas escalables, capaces de procesar grandes volúmenes de información e incorporar servicios en la nube o sistemas distribuidos cuando se requiera (Zhang et al., 2018). Esa tarea se complementa con buenas prácticas de ingeniería: integración y entrega continua, metodologías de MLOps y procesos que garanticen soluciones confiables y fáciles de mantener en el tiempo (Amershi et al., 2019).

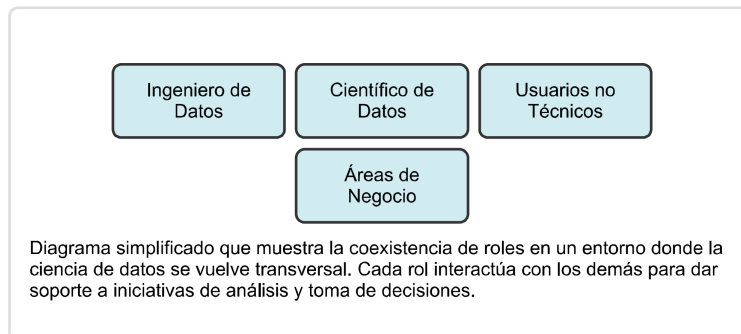


Figura 3 Mapa de roles en el ecosistema.

4.2 Aseguramiento de la calidad de datos y modelos

Cuando los datos llegan de muchas fuentes y en formatos distintos, el riesgo de errores aumenta. El ingeniero en ciencia de datos tiene que poner orden: definir reglas de validación, limpiar lo que no sirve y crear herramientas que mantengan la coherencia de la información (Kandel et al., 2018). Además de los datos, también debe vigilar los modelos. Eso significa revisar su rendimiento, ajustar parámetros y cambiar variables cuando sea necesario, para que sigan siendo útiles y precisos en el tiempo (Sculley et al., 2015).

4.3 Difusión de buenas prácticas y formación

La apertura de la ciencia de datos ha traído a la mesa a usuarios con niveles de conocimiento muy diferentes. Por eso, el ingeniero también actúa como guía y formador: organiza sesiones de capacitación, escribe materiales claros y promueve una cultura basada en buenas prácticas de estadística, programación y uso ético de la información (Cao, 2017). Al mismo tiempo se convierte en un puente entre los equipos técnicos y las áreas de negocio, asegurando que las soluciones analíticas resuelvan problemas reales y estén alineadas con los objetivos de la organización (González & Chen, 2021).

4.4 Ética y control de sesgos algorítmicos

La expansión del acceso a la analítica avanzada trae consigo un riesgo evidente: los sesgos en los modelos y el uso indebido de los datos. Frente a esto, el ingeniero debe implementar auditorías y revisiones que permitan detectar fallos y corregir sesgos en los algoritmos. También es clave aplicar técnicas de interpretabilidad como LIME o SHAP, que ayudan a entender cómo toman decisiones los modelos y a identificar diferencias de rendimiento entre distintos grupos (Amershi et al., 2019). Este esfuerzo no se hace en solitario: requiere trabajo conjunto con equipos legales y de cumplimiento, para garantizar que se respeten marcos regulatorios como el GDPR o las leyes nacionales de protección de datos (Pasquale, 2015).

5. Desafíos y oportunidades en la evolución del rol

La apertura de la ciencia de datos trae consigo un escenario doble de retos y oportunidades nuevas. Continuamente aparecen retos complejos y, al mismo tiempo, nuevas oportunidades para el ingeniero en cuanto al desarrollo profesional. Uno de los puntos más evidentes es la gestión de la tecnología. Las plataformas no-code hacen que cualquiera pueda armar un modelo inicial, pero asegurar que funcione a gran escala y que se integre en arquitecturas distribuidas sigue siendo responsabilidad de profesionales con experiencia (Sculley et al., 2015).

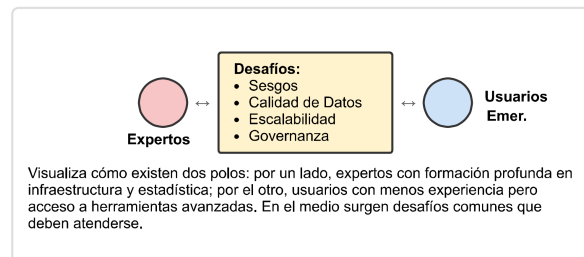


Figura 4 Brecha de conocimientos.

Para el ingeniero, esto abre la puerta a profundizar en campos como MLOps, DevOps y big data. Su papel se consolida como el de un diseñador de soluciones sólidas, capaces de adaptarse a empresas

de distintos tamaños y necesidades. A esto se suma un reto muy concreto como lo es el ciclo de vida de los modelos. Cuando hay demasiados modelos en producción sin seguimiento, la obsolescencia llega rápido. Aquí el ingeniero tiene la oportunidad de liderar procesos de actualización, reentrenamiento y validación continua, incorporando además una visión de sostenibilidad en la práctica diaria (Amershi et al., 2019).

Otro frente importante es la formación. El interés creciente de más usuarios en la analítica de datos abre un espacio para que el ingeniero actúe también como formador: diseñar capacitaciones, elaborar guías accesibles y promover una cultura de buenas prácticas que incluya la ética en el manejo de la información (Kandel et al., 2018).

Y, de manera transversal, está el tema de la ética y la responsabilidad. El acceso masivo a datos sensibles obliga a reforzar la transparencia en los algoritmos, proteger la privacidad y asegurar que los resultados sean equitativos. En este terreno, el ingeniero se convierte en garante de la integridad de los modelos, trabajando de la mano con especialistas legales y de cumplimiento normativo (Pasquale, 2015).

Factor	No-code / Low-code	Desarrollo Tradicional
Velocidad de prototipado	Rápida	Media o Baja
Requerimiento técnico	Bajo	Alto
Escalabilidad	Limitada	Alta
Flexibilidad	Baja	Alta
Riesgo de sesgo/error	Alto	Depende de la experiencia

Tabla comparativa de las principales diferencias al abordar proyectos de ciencia de datos con enfoques no-code/low-code frente a un desarrollo tradicional que requiere más programación, pero ofrece mayor control.

Figura 5 Comparación de enfoques.

Conclusiones

La democratización de la ciencia de datos se ha visto impulsada por tres grandes fuerzas como son las herramientas no-code, las plataformas de AutoML y la caída en los costos de la infraestructura tecnológica. Gracias a esa combinación, hoy resulta posible aprovechar la información no solo en grandes corporaciones, sino también en pequeñas empresas e incluso en proyectos individuales. Sin embargo, este avance no está libre de problemas y aparecen continuamente retos sobre la calidad de los datos, la gobernanza y la necesidad de aplicar buenas prácticas en todo el proceso como políticas de privacidad.

En este escenario, el ingeniero en ciencia de datos no pierde relevancia, más bien amplía sus responsabilidades. Tiene que diseñar arquitecturas que puedan escalar, vigilar que los modelos se mantengan precisos y confiables, y acompañar a equipos con poca experiencia técnica en su formación. También actúa como promotor de una cultura responsable en el uso de la información, recordando la importancia de la transparencia en los algoritmos y de reducir sesgos que pueden afectar los resultados.

Hablar de democratización no significa únicamente que existan herramientas fáciles de usar. El valor real aparece cuando se generan espacios de colaboración entre ingenieros, especialistas en datos y áreas de negocio. Esa interacción ayuda a construir organizaciones que toman decisiones con base en evidencia y con principios éticos sólidos como respaldo. De cara al futuro, es probable que ganen peso los llamados *citizen data scientists*: profesionales que, sin ser expertos en el área, aplican técnicas analíticas en su trabajo diario. Frente a este panorama, el ingeniero se consolidará como figura clave para garantizar coherencia, sostenibilidad y calidad en los proyectos de análisis de datos.

Referencias

- Amershi, S., Begel, A., DeLine, R., Gall, H., Kamar, E., Nagappan, N., ... & Zimmermann, T. (2019). Software Engineering for Machine Learning: A Case Study. In 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP) (pp. 291-300).
- Cao, L. (2017). Data Science: A Comprehensive Overview. *ACM Computing Surveys*, 50(3), 1-42.
- Dhar, V. (2013). Data Science and prediction. *Communications of the ACM*, 56(12), 64-73.
- González, D. & Chen, A. (2021). Democratizing Analytics: Low-code and No-code Approaches in Modern Enterprises. In *Proceedings of the 2021 IEEE International Conference on Big Data (Big Data)* (pp. 232-240).
- He, X., Zhao, K., & Chu, X. (2021). AutoML: A Survey of the State-of-the-Art. *Knowledge-Based Systems*, 212, 106622.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- Kandel, S., Heer, J., Plaisant, C., & Kennedy, J. (2018). Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, 17(4), 287-296.
- McKinney, W. (2017). *Python for Data Analysis* (2nd ed.). O'Reilly Media.
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.
- Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., ... & Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. In *Advances in Neural Information Processing Systems (NeurIPS)* (pp. 2503-2511).
- Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, 42, 146-157.

Capítulo 8

Sistema de interfaz cerebro computadora basado en EEG en el contexto de Internet de las cosas y ciencia de datos – Una revisión

Regina Monserrat Avelino Farfán, Carlos Leopoldo Carreón Díaz de León, Daniel
Marcelo González Arriaga

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de
Puebla, Mexico

`regina.avelinof@alumno.buap.mx`, `carlos.carreon@correo.buap.m`,
`daniel.gonzalezarr@correo.buap.mx`

Resumen.

Existe la percepción de que las máquinas sustituirán a los seres humanos, la realidad es que la intuición humana destaca al procesar información compleja. Investigaciones recientes sobre ICC han comprobado la traducción de intenciones motoras originadas en el cerebro a acciones para controlar dispositivos del Internet de las Cosas. Las EEG ofrecen una forma asequible, segura, portable, etc., de obtener dichos datos, la investigación sobre diademas electroencefalográficas inalámbricas lo demuestran. La mayor parte de la literatura aborda independientemente estos conceptos, Internet de las Cosas, ciencia de datos o sistemas ICC basados en EEG; o incluso solo las ramas del Internet de las Cosas (IIoT o IoMT). Esta revisión muestra las aplicaciones del Internet de las Cosas en diferentes sectores, la asociación entre conceptos y sus métodos por fase (adquisición de datos, procesamiento de datos, extracción de características, clasificación, control o etapa de salida), concluyendo con una explicación de su funcionamiento conjunto.

Palabras Clave: EEG, BCI, Data, Data Science, Classification, IoT.

1 Introducción

Hoy en día nuestra interacción con el mundo y el internet ha ido cambiando con la integración del Internet de las Cosas (IoT) a nuestra vida cotidiana, estos dispositivos interconectados entre si extraen información del entorno para que a partir de ella y su procesamiento se tomen decisiones en áreas como domótica, manufactura e industrias, transporte y sistemas de monitoreo de salud basados en Internet de las Cosas, a su vez de satisfacer con un bajo coste las demandas individualizadas de los clientes también llamado *personalización masiva* Neumman et al (2021). Esto último, más el devenir de

chips de computadora de bajo coste y telecomunicaciones con alto ancho de banda ha causado un auge del Internet de las Cosas estos últimos años, en palabras del McKinsey Global Institute (MGI), “el Internet de las Cosas supondrá una producción de entre 3,9 y 11,1 billones de dólares para 2025 en nueve entornos diferentes, como el comercio minorista, ciudades y fábricas”. Además, se prevé que el número de dispositivos basados en Internet de las Cosas alcance los 754,100 millones, equivalente a un aumento global de 127 dispositivos por segundo desde 2020 Ullah et al. (2025). Con este escenario económico el Internet de las cosas se posiciona entre las tecnologías de la información integradas a la industria 4.0 Neumann et al. (2021).

Por otro lado, la Interfaz Cerebro Computadora (ICC) es una arquitectura funcional y/o paradigma emergente que ha tenido avances en el control de objetos del Internet de las Cosas a través de la captura de señales cerebrales por sensores convirtiéndose así en una interfaz de entrada (input) dentro de los sistemas del Internet de las Cosas, transformando la interacción humano computadora (HCI). Aunque su enfoque inicial fue en el área de la biomedicina al aplicarse en personas con síndrome de enclaustramiento, pacientes con parálisis (cuadripléjica, paraplejia), anomalías del habla, etc., su estudio continuo permitió su extensión a áreas no medicas como realidad virtual, videojuegos, interfaz cerebro a cerebro, manejo de robots humanoides, control de silla de ruedas, etc. Yadav y Maini (2021).

En cuanto a la ciencia de datos este es un esquema o enfoque multidisciplinar en el cual se identifica, extrae y se forma una visión de los datos. Ayuda al proceso, gestión, análisis y asimilación de grandes volúmenes de datos fragmentados, estructurados o no estructurados. Sin embargo, estas soluciones propuestas según la aplicación se verán obstaculizadas porque dependiendo del objeto controlado como una silla de ruedas, artefactos del hogar o vehículos no tripulados conectados a internet se optará por métodos diferentes que se ajusten al problema, además de que por naturaleza las ICC tienen limitantes en su desempeño por factores fisiológicos Yadav y Maini (2021).

Dado este contexto de como estas tecnologías se relacionan, además de una revisión en la literatura que arroje resultados individuales en las áreas del Internet de las Cosas, ICC basado en EEG, y Data Science y una intersección parcial entre dos de las tres tecnologías expuestas, se tiene como objetivo presentar integralmente estas áreas además de mencionar los métodos que la ciencia de datos proporciona para el procesamiento de datos en algunas de las aplicaciones del Internet de las cosas controladas por ICC basados en EEG.

El trabajo se estructura como sigue: Sección 2 explicación de la metodología de búsqueda. Sección 3 descripción del estado del arte de estas tres tecnologías convergentes. Sección 4 Conclusiones y posibles mejoras que se podrían aportar a partir del alcance de esta revisión.

2 Búsqueda de la literatura

Esta sección describe el proceso de búsqueda y posterior recopilación de artículos relacionados a nuestra investigación, sistemas ICC basados en EEG con conexión a Internet de las cosas y el uso de ciencia de datos en su aplicación. Para este propósito

se utilizó un método de análisis documental llamado *cartografía conceptual*, que sistematiza los resultados previos de investigación sobre un tema en específico dentro de la literatura científica especializada Guzmán Rivera et al. (2020); el uso de Web of science, y una base de datos enfocada en la literatura biomédica, PubMed. Finalmente, para obtener resultados concretos se hicieron búsquedas booleanas. Teniendo un contexto de búsqueda, se introdujeron las siguientes búsquedas booleanas, en Web of science para la revisión de artículos indexados a revistas con un de alto o mediano impacto, y con una antigüedad máxima de 5 años garantizando así una mejor fiabilidad de la información.

("EEG signal processing" OR "electroencephalography signal processing") AND (features OR classification OR filtering)

("Internet of Things" OR "IoT") AND ("Electroencephalogram" OR "EEG")

("EEG signal processing" OR "electroencephalography signal processing") AND (features OR classification OR filtering) AND ("IoT")

("EEG signal processing" OR "electroencephalography signal processing")

2.1 Hallazgos

La cartografía conceptual resultante consta de 8 apartados tras nuestro análisis:

1. Noción:

Noción central: Sistema ICC basado en EEG para manejar Internet de las Cosas con Ciencia de Datos

Definición: Un sistema ICC basado en EEG, Internet de las Cosas y ciencia de datos es una plataforma compuesta que captura señales EEG mediante electroencefalografía, se transmiten a través de conexiones a Internet de las Cosas, y se procesan con algoritmos de ciencia de datos para crear interfaces inmediatas entre el cerebro humano y sistemas informáticos.

Propósito: Transformar señales cerebrales a comandos para el manejo de dispositivos del Internet de las Cosas por medio de técnicas de ciencias de datos.

Origen etimológico:

El término Interfaz Cerebro Computadora (ICC) fue acuñado en la década de 1970 por Jacques Vidal Vidal (1973)., su etimología proviene del inglés "Brain" (cerebro); "Computer" (computadora); "Interface" (interfaz). Este evolucionó desde pruebas básicas de lectura de señales cerebrales hasta los sistemas más recientes conformados por interpretación de electroencefalografías (EEG), conectividad a Internet de las Cosas y análisis avanzado de datos.

La electroencefalografía encuentra sus raíces en el griego "elektron" (ámbar/electricidad); "enkephalos" (cerebro); "graphein" (escribir). Fue desarrollada por Hans Berger en 1924 Yadav y Maini (2021). El Internet de las Cosas (IoT) fue conceptualizado por Kevin Ashton en 1999 Shammar y Zahary (2020)., mientras que "Ciencia de Datos" se registró por primera vez en el prefacio del libro de Naur de 1974 Ullah et al. (2025) .

2. Categoría

Categoría principal: Los sistemas ICC basados en EEG, Internet de las Cosas y ciencia de datos pueden considerarse dentro de la categoría de sistemas ciberfísicos biomédicos, que incluye tecnologías de interfaz humano computadoras avanzadas. A su vez obedece a una jerarquía más alta, el concepto de bioingeniería Terán Pérez (2016)., que engloba técnicas, procesos y tecnologías que integran sistemas biológicos con infraestructuras computacionales, electrónicas y de comunicación.

Categoría temporal: El concepto está dentro de la clasificación de tecnologías convergentes del siglo XXI, el sistema BCI-EEG- Internet de las Cosas -Ciencia de datos es un componente cuya aplicación es una de las que se espera de las mismas, crear interfaces que transfieran información del cerebro a una computadora o red Posadas Velázquez (2024). Por último, podemos categorizar a las interfaces BCI basados en EEG, Internet de las Cosas y ciencia de datos dentro de la revolución industrial 4.0 que surge de la convergencia tecnológica de diversos campos: biotecnología, cibernética, el internet, etc., Neumann et al. (2021); Posadas Velázquez (2024).

3. **Caracterización**

Características

- Captura no invasiva de señales electroencefalográficas en tiempo real.
- Implementa conectividad inalámbrica para transmisión de datos cerebrales.
- Ofrece procesamiento inteligente mediante algoritmos de machine learning.
- Funciona sobre arquitecturas distribuidas desde edge hasta cloud computing Ullah et al. (2025).
- Registra y analiza patrones neurofisiológicos ininterrumpidamente. Tiene la posibilidad de funcionar como dispositivos portables o sistemas clínicos integrados Kumar et al. (2023).
- Gestiona grandes volúmenes de datos temporales con alta frecuencia de muestreo Yadav y Maini (2021).
- Proporciona capacidades de análisis predictivo y clasificación de estados mentales. Tiene la capacidad de automatizar la detección de anomalías neurológicas

4. **Diferenciación**

Tabla 1. Ejes principales de la cartografía conceptual

Aspecto	BCI-EEG- Internet de las Cosas - Data Science	EEG Clínico Tradicional
Definición	Sistema que une interfaz cerebro-computadora con conexión a Internet de las Cosas y análisis de datos.	Sistema de monitoreo neurológico que registra actividad eléctrica cerebral para una valoración médica.
Función principal	Establecer interfaces directas cerebro-computadora con comunicación y control.	Diagnosticar condiciones neurológicas y monitoreo programado de actividad cerebral
Alcance	Preventivo, predictivo e interactivo, Enfocado en aplicaciones en tiempo real. Puede ser Wireless y portable.	Enfocado en evaluación médica. Difícil de usar sin especialistas y de difícil transporte.
Tareas clave	Clasificación de señales cerebrales para distinguir intenciones mentales Control de dispositivos externos	Diagnóstico diferencial Registro de ondas cerebrales Identificación de patologías

	Monitoreo remoto constante	Valoración de capacidad cerebral
	Predicción de eventos neurológicos	
Modo de operación	Siempre conectado, procesamiento automatizado con respuesta inmediata	Sesiones agendadas y una interpretación especializada de los resultados.
Audiencia objetivo	Pacientes con discapacidades, usuarios de tecnologías asistenciales, Utilidad para rehabilitación	Pacientes con condiciones neurológicas adversas, centros hospitalarios, clínicas especializadas
Nivel de actuación	Desde procesamiento local (edge computing) hasta análisis cloud con integración a Internet de las Cosas	Principalmente nivel hospitalario con equipos especializados
Ejemplo actividades	Transformación de datos a comandos para el control de silla de ruedas.	Detección de afecciones neurológicas, esto con la debida atención de especialistas.

5. Subdivisión

- *Por aplicación médica:* aumentativa y alternativa, Interfaces cerebro-computadora para la comunicación, Plataformas de monitoreo de epilepsia, Dispositivos de neurofeedback terapéutico, Sistemas de rehabilitación neurológica
- *Por arquitectura de procesamiento:* Sistemas edge-based, Arquitecturas cloud-native, Híbridos fog-computing, Sistemas embedded
- *Por tipo de conectividad IoT:* Redes personales (PAN) o tecnologías inalámbricas de corto alcance - Bluetooth, Zigbee, Redes locales (LAN) - WiFi, Ethernet, Redes extendidas o de área extensa (WAN) - Celular, LoRaWAN, Redes satelitales para ubicaciones remotas
- *Paradigma-Tipos de potenciales evocados:* VEP – Visual Evoked Potential: Steady-State Visual Evoked Potentials (SSVEP), Compressed-Perceptual Visual Evoked Potentials (CPVEPs), ERP Event-Related Potentials: Potenciales cognitivos: N100, P200, N200, P300, Somatosensory Evoked Potential SSEP, Event-Related Synchronization (ERS), Event-Related Desynchronization (ERD). Potenciales Corticales Lentos (SCP)
- *Por tipo de controlador:* Open-loop, machine-in-the-loop, human-in-the-loop, co-adaptive, HMCL (closed-loop)

6. Vinculación

- *Corrientes epistemológicas y técnicas:* Neurociencia: Aplicación de la física, estadística, el análisis matemático asintótico y la creación de software especializado para el análisis de datos cerebrales y crear interfaces directas Rojas et al (s.f).
- *Movimientos sociales y políticos:* Derechos de personas con discapacidad: Los sistemas ICC facilitan la adaptación de personas con capacidades limitadas a la sociedad. Protección de neuro datos personales: Según la Agencia Española de Protección de Datos (2023), es importante la protección de datos sensibles que podrían inferir en estados emocionales y cognitivos. Transhumanismo: aumento de la capacidad humana de la fusión de los mundos físico, digital y biológico Posadas Velázquez (2024).

- *Líneas investigativas:* Neurotecnología ética: Desarrollo de marcos éticos para un uso consciente de las interfaces cerebro-computadora ICC. Brain-to-brain communication: Investigación para una comunicación directa entre cerebros humanos por ICC. Yadav y Maini (2021).

7. Ejemplificación

El artículo CCA-Based Compressive Sensing for SSVEP-Based Brain-Computer Interfaces to Command a Robotic Wheelchair Rivera Flor et al. (2022)., Con el propósito de ayudar a personas con discapacidad motora propone una interfaz cerebro computadora que usa potenciales evocados visuales de estado estable (SSEVEP) y EEG para el control remoto de una silla de ruedas robótica. Para esto se utilizó un kit portable con EEG, y un mini PC (Intel NUC) que recibe los comandos, un framework que estandariza la comunicación, conexión y control entre diferentes sillas de ruedas; el Sistema de Operación Robotico (ROS) utilizando comunicación TCP/IP. El sistema inicia en el nodo sensor (diadema EEG), las señales EEG son preprocesadas aplicando referencia media común (CAR), y después filtradas por paso de banda en un rango de frecuencia de 7 a 50 Hz. A continuación, se aplica Detección Compresiva (CS) donde se genera una matriz de medición ortonormal Φ . Se multiplica por el EEG filtrado para obtener una versión comprimida y se transmite la señal sin reconstrucción. El artículo propone dos tipos de ICC basados en SSVEP:

- a. No calibrado: un análisis de correlación canónica y una matriz de compresión variando formas de onda sinusoidales y cosenoidales sintetizadas que oscilan a la frecuencia del estímulo visual.
- b. Con etapa de calibración: integra una etapa de entrenamiento con datos EEG comprimidos específicos del sujeto mejorando la precisión del ICC. Las señales SSVEP de entrenamiento se obtienen promediando múltiples ensayos. Se obtienen los coeficientes de correlación entre las épocas de EEG proyectadas de este conjunto y las señales de referencia sintetizadas basadas en patrones de onda sinusoidales y cosenoidales de la frecuencia fundamental (f_0) y armónicos que representan los estímulos visuales (u objetivos). Después, el resultado de la clasificación se asocia a un objetivo con un coeficiente de correlación máximo. La etapa de calibración usa los datos de EEG comprimidos, el modelo de clasificación puede ampliarse a nuevos datos comprimidos sin transformación adicional, lo que reduce el costo computacional. Después de la transmisión de los datos comprimidos a través de un nodo ROS se aplica un clasificador SSVEP. Ambos casos combinan la detección compresiva (CS) y el análisis de correlación canónica (CCA) para clasificar comandos. El comando recibido por el módulo de control es enviado a través del nodo al simulador de silla de ruedas desarrollado con Unity SimCardRom, que permite transmitir comandos de control a una silla de ruedas virtual o real. Para el control se definieron 40 estímulos con frecuencias entre 8 y 15.8 Hz. De los cuales se agruparon en 4 comandos: 8–9.8 Hz Retroceso; 10–11.8 Hz Derecha; 12–13.8 Hz Izquierda; 14–15.8 Hz Avanzar. Finalmente se mide su performance con ACC, ITR, TFS, TCP y latencia.

8. Metodología

1. *Fase 1: Análisis y Diseño del Sistema.* Buscar y elegir hardware EEG apropiado (resolución, portabilidad, canales). Planeación de la arquitectura

IoT (topología de red, protocolos de comunicación). Definición del pipeline completo en el procesamiento de datos y algoritmos de ML.

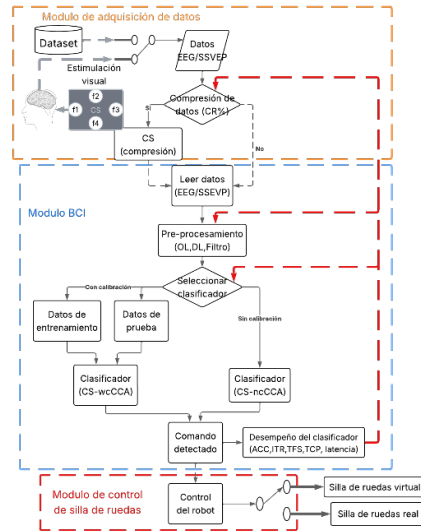


Figura 2. Diagrama de bloque del ICC propuesto basado en CS y SSVEP

2. *Fase 2: Aplicación de Infraestructura.* Ajuste de dispositivos EEG y calibración individual (depende el caso). Despliegue de red IoT (nodos ROT). Aplicación de sistemas para procesamiento local.
3. *Fase 3: Desarrollo de Algoritmos de Procesamiento.* Preprocesamiento de señales EEG (filtrado, eliminación de artefactos). Extracción de características relevantes (potenciales evocados). Entrenamiento de clasificadores
4. *Fase 4 Evaluación y pruebas.* Pruebas con múltiples configuraciones (Tasa de compresión y Longitud de la ventana de tiempo). Análisis estadístico (ANOVA). Prueba en simulador.

3 Discusión

Las diademas EEG en el contexto de la ciencia de datos actúa como una tecnología de recopilación de datos de tipo Wireless que permite al usuario realizar actividades con las manos libres Kocaoglu. (2024). En IoT la manera general en que funciona una ICC es que, mediante un equipo de recolección de señales cerebrales, las señales sin procesar se envían remotamente a través de cualquier tipo de conexión IoT a un servidor, después el servidor utiliza algún modelo preentrenado y dependiente de la persona para analizarlas. La interpretación se realiza mediante los componentes ICC adquisición de señales, pre-procesamiento, extracción de características, clasificación, e interfaz de aplicación que los transforman en comandos de control para un dispositivo IoT. Figura 3. En la adquisición de señales según las técnicas que ofrecen los ICC se clasifican como

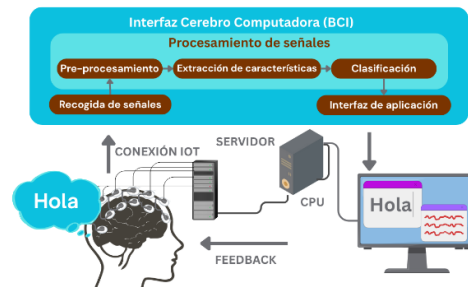


Figura 3. ICC para controlar dispositivos IoT

Invasivos: introducen electrodos en el cerebro por cirugía y aunque las señales adquiridas son de alta calidad las desventajas es que es riesgoso y no se pueden extraer señales en otras partes del cerebro más que en su lugar de implementación.

Seminvasivos: los electrodos se colocan en la parte expuesta del cerebro. Por ejemplo, la electrocorticografía (ECoG).

No invasivos: se colocan sobre el cuero cabelludo y son más seguras, sin embargo, las señales obtenidas pueden ser de más baja calidad que una invasiva. Ejemplo, la magnetoencefalografía (MEG), electroencefalografía (EEG), resonancia magnética funcional (RMF) y Espectroscopía de infrarrojo cercano (NIRS).

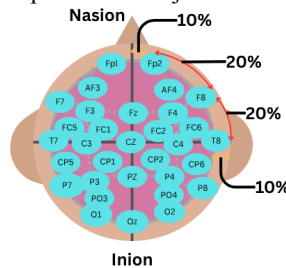


Figura 4. Sistema 10-20 para colocar diademas

El EEG tiene las ventajas de tener una Baja relación señal-ruido (SNR), una alta resolución, una rápida respuesta (0.5-130 ms) y muchos tipos en el mercado

(Emotive, Open BCI y National Instruments). En los EEG su patrón de onda varía según la posición del electrodo. Normalmente para saber su colocación se utiliza el sistema 10-20, Figura 4. El lóbulo a cada lado se representa con una letra, mientras que el hemisferio con un número. La ubicación de lóbulos frontal, parietal, occipital y temporal se denotan con F, P, O y T en este sistema. La ubicación de los electrodos en el hemisferio derecho se indica con números pares 2, 4, 6 y 8, mientras lo del izquierdo con impares 1,3,5,7.

En el preprocesamiento se extrae el ruido de la señal EEG registrada. En esta etapa se suelen registrar algunas señales eléctricas y musculares no deseadas llamadas artefactos. Pueden ser endógenas, provenientes del cuerpo o exógenas provenientes del ambiente y se eliminan antes de seguir con el procesamiento de señales para una mejor calidad de señal, un buen rendimiento de extracción de características y clasificación. Esto se realiza con filtros que en el caso de la EEG suelen estar por encima o por debajo de la frecuencia en que operan sus señales que suele ser entre 0,2 y 40 Hz. Algunos ejemplos son la Transformada Discreta de Fourier (DFT) y Filtros de Respuesta de Impulso Infinito (FIR). Algunos métodos estándar para eliminar artefactos son el filtrado lineal el análisis de componentes principales, el análisis de componentes independientes, etc.

En la extracción de características obtenemos valores significativos llamados características. Algunas técnicas para extraerlas son análisis de componentes principales, análisis de componentes independientes y Modelo Autorregresivo (AR), algoritmo genético, y selección secuencial. Se sugiere el análisis de componentes principales para extraer actividades y obtener información de la señal. El modelado de RA y algoritmos genéticos se utilizan para el reconocimiento de patrones y la extracción de características en epilépticos. Al analizar el dominio temporal, las amplitudes de señal varían con el tiempo. En esta fase tomamos una ventana temporal específica y se extraen los elementos importantes. Las técnicas orientadas a la aplicación son modelo autorregresivo AR, Codificación Predictiva Lineal LPC y Modelado Basado en núcleos KB. Después se comprimen dimensiones.

Al extraer características se definen los Potenciales Relacionados con Eventos ERDS, es decir, señales aumentadas en una banda específica: Delta (0.5-4 Hz): sueño profundo o inconsciencia. Theta (4-7 Hz): somnolencia, creatividad y entusiasmo. Alpha (7-12 Hz): relajación, calma y ojos cerrados. Mu (8-13 Hz): actividad motora. Beta (12-30 Hz): estados de alerta, concentración, cognición activa y ansiedad. Gamma (más de 30 Hz): anomalías, Agitación y memoria a corto plazo para emparejar objetos.

Además de los potenciales evocados que son respuestas eléctricas a un estímulo periódico y repetitivo registradas en un lugar puntual del sistema nervioso. Pueden ser del tipo visual, sensorial o auditivo.

En la clasificación se convierte la característica a un comando de control. Comúnmente los algoritmos de clasificación se dividen en dos categorías: lineales y no lineales. Cuando hay una relación entre inputs y outputs se utilizan técnicas lineales como El Análisis Discriminante Lineal (LDA) y la Máquina de Vectores de Soporte (SVM). Las no lineales se utilizan cuando no hay relación directa entre los patrones de entrada y salida. Ejemplos de esto son la red neuronal artificial (ANN), clasificador K-nearest neighbor y SVM's.

Con esto explicado tenemos claro que la ciencia de datos se utiliza a lo largo de una ICC, sin embargo, su aplicación se extiende más al controlar un IoT: mejorar el rendimiento a través de conocimiento. Para ello la ciencia de datos ofrece cuatro tipos de análisis de datos para determinar la mejor técnica de análisis según el contexto:

Análisis descriptivo: Examina los datos y que está ocurriendo en su contexto, para ello emplea gráficos visuales de datos como tablas, gráficos de líneas, diagramas de barra, etc.

Análisis de diagnóstico: Analiza el por qué de eventos o patrones en los datos. Para categorizarlos se utilizan los métodos Drill-down, Minería de datos, Data discovery, y correlaciones. Estos trabajan con datasets que son manipulados y pasan por varias transformaciones.

Análisis predictivo: Utiliza datos preliminares para predecir patrones de datos que puedan ser reproducibles en el futuro. Para su clasificación se usa Machine Learning, Pattern Matching, Forecasting, y el modelado predictivo.

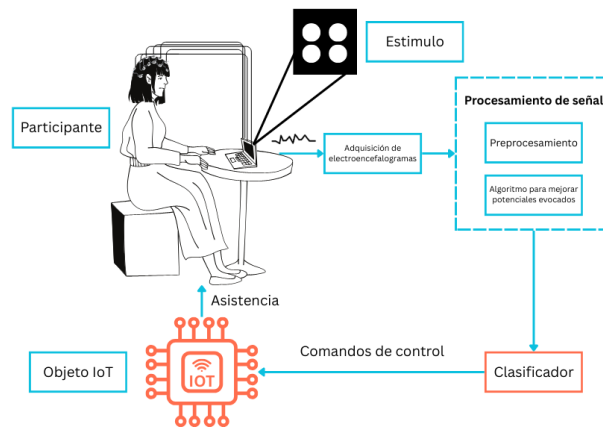


Figura 5. Diagrama unificado para manejar Internet de las cosas mediante electroencefalogramas

4 Conclusiones

Esta investigación trato de generalizar e integrar lo máximo posible como es la intersección de estas tres tecnologías convergentes y su relevancia en diversas áreas, abriéndonos camino a líneas investigativas hacia otras disciplinas, no necesariamente de ciencias exactas. Además de darnos un panorama de como diferentes técnicas dentro de la ciencia de datos nos pueden ayudar para este tipo de aplicaciones que ya son una realidad. Para esta revisión como observación se añadiría un análisis cuantitativo en el

proceso de búsqueda de artículos que hablen sobre la integración de estas tres tecnologías para filtrar calculadamente la información.

Referencias

- Abdellatif, A. A., Khafagy, M. G., Mohamed, A., & Chiasserini, C. F. (2018). EEG-Based transceiver design with data decomposition for healthcare IoT applications. *IEEE Internet of Things Journal*, 5(5).
- Afrin, S., Rafa, S. J., Kabir, M., & Gandomi, A. H. (2025). Industrial internet of things: Implementations, challenges, and potential solutions across various industries. *Computers in Industry*, 170, 104317.
- Agencia Española de Protección de Datos. (2023). Neurodatos y neurotecnología: Privacidad y protección de datos personales. <https://www.aepd.es/prensa-y-comunicacion/blog/neurodatos-y-neurotecnologia-privacidad-y-proteccion-de-datos-personales>
- Ajmeria, R., Mondal, M., Banerjee, R., Halder, T., Deb, P. K., Mishra, D., Nayak, P., Misra, S., Pal, S. K., & Chakravarty, D. (2023). A critical survey of EEG-based BCI systems for applications in industrial internet of things. *IEEE Communications Surveys & Tutorials*, 25(1).
- Du, Y., Liu, J., Wang, X., & Wang, P. (2022). SSVEP-based emotion recognition for IoT via multiobjective neural architecture search. *IEEE Internet of Things Journal*, 9(21), 21432–21443.
- Elmoussalami, H., Hui, F. K. P., & Aye, L. (2025). Electroencephalography (EEG) for psychological hazards and mental health in construction safety automation: Algorithmic systematic review (ASR). *Automation in Construction*, 170, 104317.
- Gao, X., Wang, Y., Chen, X., & Gao, S. (2021). Interface, interaction, and intelligence in generalized brain–computer interfaces. *Trends in Cognitive Sciences*, 25(8).
- Gouda, S. K., Choudhry, A., Satpathy, S. P., Shukla, K. M., Dash, A. K., & Pasayat, A. K. (2025). Integration of EEG-based BCI technology in IoT enabled smart home environment: An in-depth comparative analysis on human-computer interaction techniques. *Expert Systems with Applications*.
- Guzmán Rivera, M. Á., Escudero-Nahón, A., & Canchola-Magdaleno, S. L. (2020). Gamificación de la enseñanza para ciencia, tecnología, ingeniería y matemáticas: Cartografía conceptual. *Sinética*, (54), ene.–jun. 2020.
- Kim, M., Kim, M., Hwang, M., Kim, H., Cho, J., & Kim, S. (2019). Online home appliance control using EEG-based brain–computer interfaces. *Electronics*, 8(10), 1101. <https://doi.org/10.3390/electronics8101101>
- Klaib, A. F., Alsrehin, N. O., Melhem, W. Y., Bashtawi, H. O., & Magableh, A. A. (2021). Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and internet of things technologies. *Expert Systems with Applications*, 166, 114037.
- Klaib, A. F., Alsrehin, N. O., Melhem, W. Y., Bashtawi, H. O., & Magableh, A. A. (2025). The history, current state and future possibilities of the non-invasive brain-computer interfaces. *Medicine in Novel Technology and Devices*, 25, 100353.
- Kocaoglu, B. (2024). Data science. In *Logistics information systems* (Springer texts in business and economics). Springer, Cham. https://doi.org/10.1007/978-3-031-60290-0_8
- Kumar, A., Chakravarthy, S., Nanthamornphong, A., Craik, A., González-España, J. J., Alamir, A., Edquiang, D., Wong, S., Sánchez Rodríguez, F., Feng, J., Gerard, F., & Contreras-Vidal, J. L. (2023). Design and validation of a low-cost mobile EEG-based brain–computer interface. *Sensors*, 23, 5930.
- Kumar, A., Chakravarthy, S., & Nanthamornphong, A. (2023). Energy-efficient deep neural networks for EEG signal noise reduction in next-generation green wireless networks and industrial IoT applications. *Symmetry*, 15(12), 2129.

- Liu, S., Ming, Z., Liu, M., Zhang, D., Liu, Z., Chen, Q., Ma, L., Luo, J., Suo, D., Zhang, J., & Yan, T. (2024). Remote-oriented brain-controlled unmanned aerial vehicle for IoT. *IEEE Internet of Things Journal*, 11(17), 29202–29214. <https://doi.org/10.1109/jiot.2024.3406837>
- Neumann, W. P., Winkelhaus, S., Grosse, E. H., & Glock, C. H. (2021). Industry 4.0 and the human factor – A systems framework and analysis methodology for successful development. *International Journal of Production Economics*, 233, 107992.
- Niedermeyer, E., & da Silva, F. L. (2004). Electroencephalography (EEG): Basic principles, clinical applications, and related fields. *Clinical Neurophysiology*, 115, 948–960.
- Posadas Velázquez, Y. (2024). Las tecnologías convergentes del siglo XXI ¿Camino hacia el transhumanismo? *Eutopía*, 13(35), 79–88. <https://www.revistas.unam.mx/index.php/eutopia/article/view/88051>
- Ranjan, R., Rana, O., Nepal, S., Yousif, M., James, P., Wen, Z., Barr, S., Watson, P., Jayaraman, P. P., Georgakopoulos, D., Villari, M., Fazio, M., Garg, S., Buyya, R., Wang, L., Zomaya, A. Y., & Dustdar, S. (2018). The next grand challenges: Integrating the internet of things and data science. *IEEE Cloud Computing*, 5(3), 12–26.
- Rivera-Flor, H., Gurve, D., Floriano, A. S. da P., Delisle-Rodriguez, D., Mello, R., & Bastos-Filho, T. F. (2022). CCA-based compressive sensing for SSVEP-based brain–computer interfaces to command a robotic wheelchair. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–10.
- Rojas Leguizamón, M., Molina Del Rio, J., Ramírez Flores, A. G., Hidalgo Aguirre, R. M., & Figueroa Jiménez, M. D. (n.d.). Una mirada a las neurociencias: De las neuronas a la cognición (pp. 69–70). Universidad de Guadalajara. <https://www.researchgate.net/publication/355048943>
- Shammar, E. A., & Zahary, A. T. (2020). The internet of things (IoT): A survey of techniques, operating systems, and trends. *Library Hi Tech*, 38(1), 5–66.
- Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. M. Z., Paul, R., Smriti, K., Naik, N., & Somani, B. K. (2021). The role of data science in healthcare advancements: Applications, benefits, and future prospects. *Irish Journal of Medical Science*, 191(4), 1473–1483.
- Terán Pérez, D. M. (2016). *Bioingeniería*. Alfaomega Grupo Editor.
- Ullah, I., Adhikari, D., Su, X., Palmieri, F., Wu, C., & Choi, C. (2025). Integration of data science with the intelligent IoT (IIoT): Current challenges and future perspectives. *Digital Communications and Networks*, 11(13), 280–298.
- Värbu, K., Muhammad, N., & Muhammad, Y. (2022). Past, present, and future of EEG-based BCI applications. *Sensors*, 22(9), 3331. <https://doi.org/10.3390/s22093331>
- Vidal, J. J. (1973). Toward direct brain–computer communication. *Annual Review of Biophysics and Bioengineering*, 2, 157–180.
- Yadav, H., & Maini, S. (2021). Electroencephalogram based brain–computer interface: Applications, challenges, and opportunities. *Multimedia Tools and Applications*, 82, 47003–47047.
- Zhang, R., Wang, Q., Li, K., He, S., Qin, S., Feng, Z., Chen, Y., Song, P., Yang, T., Zhang, Y., Yu, Y., Hu, Y., Shao, M., & Li, Y. (2017). A BCI based environmental control system for patients with severe spinal cord injuries. *IEEE Transactions on Biomedical Engineering*, 64(8), 170–177.
- Zhang, X., Yao, L., Zhang, S., Kanhere, S. S., Sheng, Q., & Liu, Y. (2019). Internet of things meets brain–computer interface: A unified deep learning framework for enabling human–thing cognitive interactivity. *IEEE Internet of Things Journal*, 6(2), 2084–2092.

Capítulo 9

Marco de referencia para la acreditación de programas educativos de nivel superior en México afines a la Ciencia de Datos

Juan Carlos Conde-Ramírez, María Teresa Torrijos-Muñoz, Mireya Tovar-Vidal,
Jaime Alejandro Romero-Sierra

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, México

carlos.conder@correo.buap.mx, teresa.torrijos@correo.buap.mx,
mireya.tovar@correo.buap.mx, jaime.romeros@correo.buap.mx

Resumen. En este trabajo se presenta un marco de referencia para la acreditación de Programas Educativos de nivel licenciatura afines a la Ciencia de Datos, que puede servir como guía para diseñar, evaluar y mejorar programas educativos, asegurando de este modo su calidad y pertinencia a un campo en constante evolución. Se enfatizan los principales enfoques y criterios de acreditación de la educación superior en México, y se presentan a los organismos de acreditación más importantes, tanto a nivel nacional como internacional. Además, se analiza el rol de la evaluación y mejora continua, identificando a los principales indicadores de desempeño, la autoevaluación del programa y la adaptación a nuevas tendencias. Finalmente, se exploran los desafíos en la acreditación de este tipo de programas educativos como lo son: la inclusión de criterios éticos en la enseñanza, la evolución de los estándares educativos ante los avances tecnológicos y la necesidad de modelos educativos más flexibles para responder a la demanda del mercado laboral.

Palabras Clave: Acreditación, ciencia de datos, programa educativo, calidad, pertinencia, marco de referencia, educación superior, evaluación, mejora continua, mercado laboral.

1 Introducción

Según la UNESCO¹ (2025), existen aproximadamente 254 millones de estudiantes universitarios de todo el mundo. Pese a la demanda, la tasa mundial de inscripción a educación superior representa apenas el 42%, detectando además grandes desigualdades entre países y regiones. Y hace hincapié en la inclusión del reconocimiento de los diplomas y títulos, así como en la garantía de la calidad, más en países en vías de desarrollo.

¹ Organismo de las Naciones Unidas con la directriz de cooperar con todos los países para lograr una educación de calidad

Uno de los medios más importantes para garantizar la calidad en la educación es la acreditación, proceso formal mediante el cual se reconoce que un plan de estudios, perteneciente a determinada institución, cumple con ciertos estándares de calidad preestablecidos (ANUIES, s.f.).

A nivel internacional, la UNESCO emplea esfuerzos para mejorar las oportunidades de movilidad de los estudiantes, mediante la Convención Mundial sobre el Reconocimiento de las Cualificaciones relacionadas con la Educación Superior, y el Pasaporte UNESCO de cualificaciones para refugiados y migrantes vulnerables, entre varios instrumentos. Asimismo, contribuye para que todos los países desarrollen la capacidad de crear redes y organizaciones que avalen la calidad en la educación superior (UNESCO, 2025).

A nivel nacional, es evidente que el desarrollo económico de la región es un indicador relevante para identificar los perfiles profesionales que la educación superior debe tener en cuenta. Por ejemplo, de acuerdo con información del Instituto Nacional de Estadística y Geografía (INEGI) y del Gobierno del Estado de Puebla, el crecimiento económico de la región es particularmente significativo en sectores como: la manufactura, la industria automotriz, el comercio y los servicios (SPF, 2024). Vale la pena enfatizar el uso de técnicas de análisis de datos, de inteligencia artificial (IA) y de modelado predictivo para generar soluciones de software con un gran impacto social en áreas como: la movilidad urbana, el medio ambiente y la seguridad.

Por lo tanto, la creación y actualización de Programas Educativos (PE) representan una oportunidad estratégica. Formar profesionales con los conocimientos y habilidades contemporáneas contribuye a mejorar la calidad de vida de su entorno, de la región y del país. El presente trabajo se organiza de la siguiente manera: 2. Factores de acreditación en la Educación Superior a nivel Internacional, 3. Organismos acreditadores más importantes, 4. Criterios de acreditación para PE de Ciencia de Datos, 5. Evaluación y mejora continua, 6. Desafíos y tendencias en la educación y 7. Conclusiones.

2 Factores de acreditación en la Educación Superior a nivel Internacional

Como ya se mencionó, la acreditación es un proceso fundamental. Permite evaluar tanto la calidad de las instituciones educativas como la de sus programas académicos. Esto garantiza su importancia en el ámbito global, nacional y regional. La acreditación permite fomentar valores como la responsabilidad y la transparencia. Esto fortalece la confianza de la sociedad en el sistema educativo (UPlanner, 2022).

A pesar de estos avances en este ámbito, la acreditación también enfrenta principalmente dos desafíos (ANUIES, s.f.):

- a. La homogeneización de los criterios entre los distintos organismos evaluadores.
- b. Aseguramiento de procesos equitativos e inclusivos para cada institución, sin importar su tamaño o ubicación geográfica.

En seguida, se presentan los principales factores que deberían considerar los programas educativos de nivel superior durante un proceso de acreditación:

- **Metodologías para diseñar perfiles:** El uso de técnicas estructuradas facilita la definición de los conocimientos y habilidades vinculadas a cada perfil profesional (como el análisis funcional). Asegurando así su pertinencia (Greciet & Villarroel, 2020).
- **Necesidades del mercado laboral:** Estudios, como el realizado por la Universidad de Andalucía, se han encargado de analizar los perfiles profesionales con una demanda en ascenso, contribuyendo a que las propuestas educativas cumplan con las necesidades presentes y futuras del campo laboral (Universidad Internacional de Andalucía, 2020).
- **Principales Actores:** El fin es garantizar que los perfiles profesionales reflejen la situación y la demanda del mercado laboral actual. Es crucial la colaboración entre instituciones educativas, organismos certificadores, empleadores y otros actores (OIE, 2022).
- **Evaluación continua:** La validación de perfiles profesionales deben incorporar un sistema de evaluación que facilite la actualización constante. Esto con el fin de ajustarse a los cambios sociales, tecnológicos y económicos (CEEAD, 2020).
- **Superación profesional continua:** Se debe fomentar la educación continua para que los profesionales actualicen sus capacidades y competencias. Es importante preservar la pertinencia de su perfil en el ámbito laboral (Pérez, 2025).
- **Competencias internacionales:** Las certificaciones profesionales, basadas en modelos internacionales, contribuyen a la confianza de que los perfiles se ajusten a estándares de calidad mundiales. Además, esto impulsa la movilidad académica y laboral. El fin es garantizar una competitividad más justa en un mercado tan globalizado (Velasco & Aguado, 2010).

Estos elementos e indicadores garantizan que los perfiles profesionales permanezcan en sintonía con las exigencias del mercado de trabajo, fomentando la posibilidad de empleo y el crecimiento económico sostenido.

3 Organismos acreditadores más importantes

Como menciona Rama (2009):

La acreditación internacional es una derivación de la complejización de los procesos de evaluación y acreditación, y se constituye como un componente que deben asumir los sistemas de educación en el actual contexto de la globalización y la sociedad del conocimiento.

La Revista Española de Educación Comparada seleccionó en 2006 a los principales organismos internacionales vinculados con la acreditación de programas de Educación Superior; esta lista se muestra en la **Tabla 1**. Sin embargo, una lista completa y actualizada de las organizaciones reconocidas puede consultarse en (Academia Lab, 2025).

Tabla 1.*Principales organismos responsables de la acreditación a nivel internacional*

PAÍS	Organismo acreditador
AMERICA LATINA	
ARGENTINA	Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU)
BRASIL	Sistema Nacional de Reconocimiento de la Enseñanza Superior
CHILE	Consejo Superior de Educación (CSE)
COLOMBIA	Consejo Nacional de Acreditación
MÉXICO	Consejo para la Acreditación de la Educación Superior (COPAES)
OTROS	
EEUU	Comisión de Acreditación para Universidades Públicas y Privadas (WASC) (carreras técnicas) / Comisión de Instituciones de Enseñanza Superior / Asociación de Nueva Inglaterra de Escuelas y Universidades (NEASC)
CANADÁ	Comisión de Evaluación de la Enseñanza Universitaria (CEEC)
JAPÓN	Asociación de Acreditación Universitaria de Japón (JUAA)
REINO UNIDO	Agencia de Acreditación de la Calidad para la Educación Superior (QAA)
RUSIA	Centro de Acreditación Nacional de la Federación Rusa (NAC)
FRANCIA	Comité Nacional de Evaluación / Inspectorado General de Educación y Desarrollo (IGAENR)
ESPAÑA	Agencia Nacional para la Evaluación de la Calidad y la Acreditación
SUIZA	Centro de Acreditación y Aseguramiento de la Calidad de las Universidades Suizas (OAQ)
ALEMANIA	Consejo de Acreditación / Instituto para el Aseguramiento de la Calidad, la Certificación y la Acreditación (ACQUIN) / Conferencia de Rectores de Educación Superior - Aseguramiento de la calidad de proyectos

Fuente: Tomada de la *Revista Española de Educación Comparada*, 12 (2006), 81-112

En México, hoy en día la evaluación y acreditación de la educación superior se realiza por medio de organismos e instancias especializadas. Los que evalúan los PE, en sus niveles de técnico superior universitario, profesional asociado y licenciatura, deben reconocer en primer término a las propias Instituciones de Educación Superior (IES) que los imparten y que periódicamente aplican las *autoevaluaciones* correspondientes. Por otro lado, están las evaluaciones externas por pares académicos. Las realizan los Comités Interinstitucionales para la Evaluación de la Educación Superior (CIEES) y los organismos acreditadores reconocidos por el Consejo para la Acreditación de la Educación Superior (COPAES). Adicionalmente, el país cuenta con organismos e instituciones que evalúan los PE de posgrado, que son las propias IES a través de sus autoevaluaciones. Los CIEES, el Programa Nacional de Posgrados de Calidad de la Secretaría de Educación Pública (SEP) y el Consejo Nacional de Ciencia y Tecnología (CONACyT) (De la Garza Aguilar, 2013).

La COPAES² es un organismo mexicano que tiene como función principal coordinar y supervisar a los distintos comités y organismos acreditadores de PE de nivel superior en México. Su objetivo es garantizar que los procesos de acreditación se realicen bajo estándares de calidad y pertinencia, contribuyendo así a la mejora continua de la educación superior en el país. Sus funciones más importantes son (COPAES, s.f.):

- **Coordinación de organismos acreditadores:** COPAES no acredita directamente PE, sino que reconoce y avala a los organismos acreditadores especializados por áreas de conocimiento.
- **Establecimiento de criterios y estándares:** Define los lineamientos y criterios que deben seguir los organismos acreditadores para evaluar la calidad de PEs.
- **Reconocimiento oficial:** Los programas acreditados por organismos avalados por COPAES son reconocidos por la Secretaría de Educación Pública (SEP) y otras instancias educativas.
- **Promoción de la calidad educativa:** Fomenta la mejora continua de la educación superior a través de procesos de evaluación y acreditación.
- **Ámbito de acción:** Trabaja con programas de licenciatura y posgrado en diversas áreas del conocimiento, desde ingenierías hasta humanidades y ciencias sociales.
- **Transparencia y rigor:** Asegura que los procesos de acreditación sean transparentes, objetivos y basados en estándares internacionales.

Con base en lo anterior, en la **Tabla 2** se comparan las dos organizaciones mexicanas de acreditación, avaladas por la COPAES, encargadas de garantizar la calidad en PEs de nivel licenciatura en áreas de conocimiento afines a Ingeniería y Computación:

Tabla 2.

Organizaciones mexicanas acreditadoras de licenciaturas en Ingeniería y Computación.

CONAIC (Consejo Nacional de Acreditación en Informática y Computación, A.C.)	
• Ámbito de acción: Específico para programas relacionados con informática, computación y tecnologías de la información. • Enfoque: Acredita programas de licenciatura y posgrado en estas áreas, asegurando que cumplen con estándares de calidad. • Categorías de los criterios de acreditación de Programas Académicos (PA):	
<i>Dimensión I: Gestión del PA</i>	<i>Dimensión II: Recursos entorno al PA</i>
1. Plan de estudios	6. Personal académico
2. Estudiantes	7. Servicios de apoyo al aprendizaje
3. Formación integral	8. Infraestructura y equipamiento
4. Gestión y operación	<i>Dimensión III: Resultados del PA</i>
5. Sistema interno de calidad	9. Resultados de aprendizaje
	10. Resultados de gestión
CACEI (Consejo de Acreditación de la Enseñanza de la Ingeniería, A.C.)	
• Ámbito de acción: Específico para programas de ingeniería y tecnología	

² Consejo para la Acreditación de la Educación Superior, A.C.

- **Enfoque:** Acredita programas de licenciatura en ingeniería y afines, asegurando que cumplen con estándares de calidad educativa.
- **Categorías de los criterios de acreditación de Programas Académicos (PA):**

1. Estudiantes	5. Personal académico
2. Plan de estudios	6. Soporte institucional
3. Objetivos internacionales	7. Mejora continua
4. Atributos de egreso	8. Área disciplinar del programa educativo

Fuente: Tomada del Marco de Referencia del CONAIC (2024), 36-63 y del Marco de Referencia del CACEI (2025), 4-68

Los criterios específicos y su descripción, para cada una de las categorías presentadas en la **Tabla 2**, pueden ser consultados en (CONAIC, 2024) y en (CACEIC, 2024).

4 Criterios de acreditación para PE de Ciencia de Datos

Según el Foro Económico Mundial en Tuduri (2025): “la industria de la ciencia de datos creció un 650% desde 2012, y se espera que la demanda mundial de profesionales de este rubro aumente en 11.5 millones para 2026”. La demanda laboral en Ciencia de Datos se verá impulsada por la globalización y la competencia internacional, mismas que están obligando a las empresas a adoptar la ciencia de datos para mantenerse relevantes. Los factores clave que disparan la demanda laboral en Ciencia de Datos son:

- **Digitalización:** La transformación digital está en marcha en todos los sectores, lo que lleva a un aumento en la cantidad de datos generados. Las empresas necesitan especialistas capaces de gestionar y analizar estos datos para adquirir *insights* útiles. Esta transformación digital impacta tanto a pequeñas y medianas empresas como a grandes corporaciones, que aspiran a optimizar sus procesos internos.
- **Análisis avanzado y Big Data:** Actualmente muchas organizaciones destinan recursos a tecnologías de Big Data y análisis sofisticado de datos para potenciar su competitividad, lo cual demanda profesionistas capaces de analizar grandes cantidades de datos. Los progresos en las tecnologías de almacenamiento y procesamiento están posibilitando que las compañías gestionen grandes volúmenes de datos de manera rápida y eficaz, lo que hace que la ciencia de datos sea más asequible y potente.
- **Automatización e IA:** La incorporación de la IA y la automatización en los procedimientos corporativos está generando una necesidad de científicos de datos capaces de elaborar y administrar modelos predictivos. Estos modelos resultan esenciales para labores como el pronóstico de tendencias de mercado, la mejora de la logística y la personalización de la experiencia del usuario (Tuduri, 2025).

Para elegir las categorías y criterios de acreditación de PE en Ciencia de Datos de nivel licenciatura, que sean pertinentes en la situación actual, es crucial tener en cuenta los elementos previamente mencionados (digitalización, análisis avanzado y Big Data, automatización e IA). Sus fundamentos deben asegurar que los PE formen profesionales

aptos para afrontar los retos tecnológicos y científicos del mundo contemporáneo. En la **Tabla 3**, se proponen las categorías y criterios de acreditación especiales, basados en la categorías y criterios definidos originalmente por el CONAIC y el CACEI:

Tabla 3.

Categorías y criterios de acreditación especiales, propuestos para PE en Ciencia de Datos de nivel licenciatura.

Categoría	Descripción	Criterios
1. Plan de estudios	El plan de estudios debe estar alineado con las demandas del mercado laboral y las tendencias tecnológicas actuales.	1. Inclusión de materias clave como Big Data, ML, IA, Estadística Avanzada, Visualización de Datos y Gestión de Bases de Datos. 2. Flexibilidad para actualizar contenidos en función de los avances tecnológicos. 3. Integración de herramientas y tecnologías emergentes (ej.: Python, R, Hadoop, Spark, TensorFlow, etc.).
2. Estudiantes (Investigación y desarrollo)	El programa debe fomentar en el estudiante la investigación en ciencia de datos y áreas relacionadas.	1. Participación de estudiantes y docentes en proyectos de investigación. 2. Publicación de resultados en revistas científicas o conferencias especializadas. 3. Colaboración con centros de investigación y desarrollo tecnológico.
3. Formación Integral	Los estudiantes deben desarrollar habilidades prácticas que les permitan resolver problemas complejos en entornos reales.	1. Uso de casos de estudio y proyectos aplicados en colaboración con la industria. 2. Acceso a datos reales y herramientas profesionales para análisis y procesamiento. 3. Fomento de la experimentación y la innovación en el manejo de datos.
4. Sistema interno de calidad	El programa debe ser innovador y capaz de adaptarse rápidamente a los cambios tecnológicos.	1. Incorporación de metodologías y enfoques disruptivos en la enseñanza. 2. Fomento de la creatividad y el emprendimiento en el uso de datos. 3. Actualización periódica del plan de estudios para incluir nuevas tecnologías y metodologías.
5. Personal académico	Los profesores deben tener experiencia y conocimientos actualizados en ciencia de datos y áreas afines.	1. Docentes con formación académica y experiencia profesional en análisis de datos, IA o campos relacionados. 2. Participación en investigación, publicaciones o proyectos de innovación. 3. Capacitación continua en tecnologías emergentes.
6. Servicios de apoyo al aprendizaje	Los estudiantes deben comprender las implicaciones éticas y sociales del uso de datos y la inteligencia artificial	1. Inclusión de talleres o seminarios, sobre ética en ciencia de datos, privacidad y seguridad de la información. 2. Discusión de casos relacionados con el uso responsable de datos y algoritmos. 3. Promoción de la transparencia y la equidad en el análisis de datos.

7. Infraestructura y equipamiento	La institución debe contar con recursos tecnológicos avanzados para apoyar el aprendizaje y la investigación.	1. Disponibilidad de laboratorios especializados con software y hardware actualizado. 2. Acceso a plataformas en la nube para el procesamiento de grandes volúmenes de datos. 3. Conectividad y capacidad de cómputo suficientes para trabajar con Modelos de IA y Big Data.
8. Resultados de aprendizaje	El programa debe demostrar que los egresados tienen las competencias necesarias para desempeñarse en el campo laboral y científico de la ciencia de datos.	1. Implementación de mecanismos para medir el dominio de habilidades técnicas y blandas (ej.: pensamiento crítico, comunicación efectiva, trabajo en equipo). 2. Seguimiento de egresados para evaluar su inserción laboral y desempeño profesional. 3. Retroalimentación continua de empleadores y egresados para mejorar el programa.
9. Resultados de gestión	El programa debe establecer vínculos sólidos con empresas y organizaciones que utilicen ciencia de datos.	1. Colaboración con empresas para el desarrollo de proyectos conjuntos o prácticas profesionales. 2. Participación de expertos de la industria en la actualización del plan de estudios. 3. Oportunidades para que los estudiantes trabajen en problemas reales del sector productivo.
10. Objetivos internacionales	El programa debe estar alineado con estándares internacionales y fomentar la movilidad académica y profesional.	1. Colaboración con instituciones extranjeras para intercambios académicos y proyectos de investigación. 2. Adopción de estándares internacionales en ciencia de datos (por ejemplo, de la ACM o IEEE). 3. Uso de literatura y recursos en inglés.

5 Evaluación y mejora continua

La UNESCO y la ANUIES coinciden en que la mejora continua es un eje fundamental para garantizar la calidad, relevancia y equidad en la educación superior.

En (UNESCO, 2021) se promueve la mejora continua en la educación superior como parte de su agenda para el Objetivo de Desarrollo Sostenible 4 (ODS 4): "Educación de calidad". Sus principios incluyen:

- **Enfoque en la calidad:** Destaca la necesidad de sistemas de evaluación y acreditación que fomenten la mejora permanente, no solo el cumplimiento de estándares estáticos. Recomendamos la revisión de demandas sociales, tecnológicas y laborales del siglo XXI, para adaptar los modelos educativos.
- **Innovación y flexibilidad:** Fomenta la incorporación de tecnologías digitales y metodologías pedagógicas innovadoras (por ejemplo: aprendizaje híbrido, micro credenciales). Afirmamos la relevancia de la investigación aplicada y su conexión con los sectores de producción.
- **Equidad e inclusión:** Subrayamos que la mejora continua debe reducir brechas. El acceso debe existir también para grupos vulnerables (mujeres, minorías, poblaciones rurales).

Por su parte, las propuestas la ANUIES (2021) cumplen con las normas internacionales. Subrayan los desafíos locales en su "Visión y Acción - 2030":

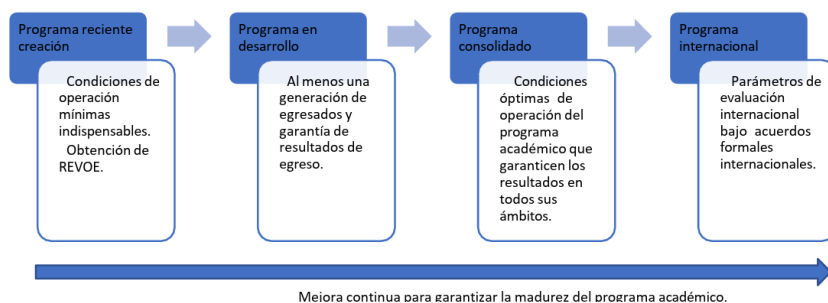
- **Evaluación formativa:** La retroalimentación a las instituciones depende de los diagnósticos regulares. Mismos sobre los que se debe apoyar la mejora continua (como evaluaciones internas o indicadores de desempeño). Se deben reprobar los procesos burocráticos de evaluación y apoyar mecanismos más ágiles.
- **Participación y Autoevaluación:** Sugiere que las instituciones fomenten una cultura de autoevaluación, que incluya a alumnos, profesores y empleadores. Modelos como los CIEES (Comités Interinstitucionales para el Análisis de la Educación Superior).
- **Vinculación con el entorno:** La mejora debe responder a necesidades regionales y nacionales, como la empleabilidad de egresados o la innovación tecnológica.

Las similitudes entre UNESCO y ANUIES, más allá de la acreditación, radican en que ambas cuestionan los sistemas de evaluación estrictos y sugieren que el proceso de mejora continua sea dinámico y contextual. Además, están de acuerdo en integrar herramientas digitales y enfoques sostenibles en los planes de estudio, sin perder de vista el enfoque humano, ya que consideran que la calidad educativa debe medirse también por su impacto social (p.ej. reducción de desigualdades).

Retornando al proceso de acreditación, un modelo de evaluación para dicho fin debe orientar la integración de un proceso escalonado. Proceso con elementos interrelacionados que definen pautas sobre qué y cómo acreditar, y que con el paso del tiempo se consolidará a través de la mejora continua (CONAIC, 2024). El modelo cuenta con una serie de etapas para lograr madurar al programa académico a lo largo de diferentes ciclos a saber. La Figura 1, muestra cómo se van integrando los diferentes niveles de consolidación de un programa académico y describe los componentes indispensables en cada uno de ellos.

Figura 1.

Modelo general de evaluación CONAIC 2023.



6 Desafíos y tendencias en la educación

De acuerdo con la UNESCO y la ANUIES algunos de los desafíos y tendencias en la acreditación de PE, para cualquier área de conocimiento, se clasifican en tres grupos: 1) la evolución de los estándares ante avances tecnológicos, 2) la inclusión de criterios éticos en la enseñanza y 3) la necesidad de modelos educativos más flexibles para responder a la demanda del mercado laboral.

Clasificación	Desafíos	Tendencias
1. Evolución de estándares y avances tecnológicos	Actualización constante: Los programas deben renovar contenidos rápidamente para incluir tecnologías emergentes (ej.: quantum computing, LLMs, edge analytics). Brecha docente: Muchos profesores no dominan herramientas avanzadas (ej.: TensorFlow 2.0, plataformas de MLOps). Infraestructura costosa: Acceso limitado a recursos como GPUs o datasets masivos en instituciones con bajos presupuestos.	Microcredenciales: Certificaciones específicas en tecnologías (ej.: AWS Data Analytics, Google Cloud AI) como complemento al currículo tradicional. Simulaciones y laboratorios virtuales: Uso de entornos cloud (ej.: Google Colab, Kaggle) para prácticas accesibles. Estándares globales: Alineación con marcos como ACM Data Science Competencies o IEEE/ISO para garantizar empleabilidad internacional.
2. Criterios éticos	Sesgos en algoritmos: Falta de formación en AI fairness y interpretabilidad en los programas. Privacidad y regulaciones: Cumplir con normativas como GDPR (UE) o LGPD (Brasil) en proyectos estudiantiles. Uso responsable de datos: Riesgo de que estudiantes apliquen técnicas sin considerar impactos sociales (ej.: deepfakes).	Ética integrada en el currículo: Cursos obligatorios sobre Responsible AI, ética de datos y gobernanza. Colaboración interdisciplinaria: Proyectos con áreas como derecho o sociología para abordar implicaciones éticas. Auditorías de algoritmos: Enseñanza de herramientas para detectar sesgos (ej.: IBM Fairness 360, Google What-If Tool).
3. Modelos educativos flexibles	Resistencia al cambio: Estructuras rígidas en universidades que dificultan la adopción de modalidades híbridas o <i>bootcamps</i>. Personalización del aprendizaje: Atender diversidad de niveles entre estudiantes (ej.: unos dominan Python y otros no).	Aprendizaje Basado en Proyectos (ABP): Vinculación con organizaciones y solución a problemas con datos reales. Rutas de aprendizaje personalizadas: A través de la acreditación de asignaturas precedentes o cursos "a la carta" (p. ej.: elegir entre Visión por Computadora o PLN).

	Validación de habilidades: Cómo certificar competencias no tradicionales (ej.: análisis de datos en proyectos reales).	Educación híbrida: Combinación de MOOCs (Coursera, edX) con mentorías presenciales.
--	--	--

7 Conclusiones

Debido a que la acreditación en la educación superior involucra un proceso dinámico, este trabajo logra ofrecer un marco de referencia que pretende servir como guía para la creación, evaluación y perfeccionamiento de los PE relacionados con la Ciencia de Datos a nivel licenciatura. Asegurando la pertinencia y calidad en un sector que está en constante cambio, así como ayudando a los empleadores a tener confianza plena en un PE que potencie las posibilidades laborales de los egresados.

Se ha comprobado que en México los procesos de acreditación han progresado significativamente. Se ha convertido en una herramienta fundamental para continuar con el mejoramiento del sistema educativo. En los últimos años, se han implementado dos acciones relevantes: 1) la creación y desarrollo de organismos especializados, y 2) el fomento de una cultura institucional de evaluación continua.

En consecuencia, es esencial fortalecer la cultura evaluativa y promover la colaboración entre entidades acreditadoras, instituciones y gobierno. Esto fomenta un sistema educativo en la educación superior de calidad que cubra las necesidades del mercado laboral y de la sociedad. Por tanto, la colaboración es fundamental para afrontar los desafíos actuales y futuros.

Por lo tanto, es evidente que la actualización o mejora de los programas educativos debe ser guiada por evaluaciones continuas. Se asegurará la conservación de altos estándares de calidad. Además, no se descuidará el enfoque sustentable y humanista de las universidades e instituciones educativas de nivel superior.

Referencias

- Academia Lab. (2025). Lista de organizaciones reconocidas de acreditación de educación superior. Recuperado el 23 de marzo de 2025, de Enciclopedia: <https://academia-lab.com/enciclopedia/lista-de-organizaciones-reconocidas-de-acreditacion-de-educacion-superior/>
- ANUIES. (2021). Visión y acción 2030: Propuesta de la ANUIES para renovar la educación superior en México. Asociación Nacional de Universidades e Instituciones de Educación Superior. Obtenido de https://coatf.cecylt9.ipn.mx/docs/Vision_accion2030.pdf
- ANUIES. (s.f.). La evaluación y acreditación de la educación superior en México. Obtenido de Publicaciones ANUIES: <http://publicaciones.anui.es.mx/acervo/revsup/res101/txt6.htm>
- CACEIC. (2024). Manual del Marco de Referencia 2025 (MR 2025) para la acreditación de programas de Ingeniería. Ciudad de México, México: Consejo de Acreditación de la Enseñanza

- de la Ingeniería, A.C. Obtenido de https://www.cacei.com.mx/wp-content/uploads/2025/02/dac_02_doc_mr_20251.pdf
- CEEAD. (2020). Orientaciones metodológicas para la elaboración de perfiles basados en estándares de competencia profesional. Centro de Estudios sobre la Enseñanza y el Aprendizaje del Derecho, A.C. Obtenido de https://cejume.mx/wp/wp-content/uploads/2020/10/201007_Orientaciones_Metodologicas_Cejume-1.pdf
- CONAIC. (2024). Marco General de Referencia para los Procesos de Acreditación de Programas Académicos de TSU, Licenciatura, Investigación y Posgrado. CDMX: Consejo Nacional de Acreditación en Informática y Computación, A.C. Obtenido de <https://www.conaic.net/publicaciones/1.%20Marco%20de%20Referencia%20integrado%20de%20CONAIC%202024.pdf>
- COPAES. (s.f.). Consejo para la Acreditación de la Educación Superior, A.C. Recuperado el 23 de marzo de 2025, de copaes.org: <http://copaes.org>
- De la Garza Aguilar, J. (2013). La evaluación de programas educativos del nivel superior en México: Avances y perspectivas. Perfiles educativos, 35(spe), 33-45. Recuperado el 20 de marzo de 2025, de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982013000500004&lng=es&tlng=es
- Greciet, P., & Villarroel, S. (2020). Metodología para la elaboración de los perfiles profesionales del Catálogo de Perfiles Profesionales (CPP). Programa EUROsociAL+. Obtenido de <https://eurosoci.al.eu/wp-content/uploads/2020/11/Metodologia-CPP-24-09-2020.pdf>
- INEGI. (2016). Estructura económica de Puebla en síntesis. (INEGI, Ed.) Instituto Nacional de Estadística y Geografía. Obtenido de https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/producto_s/nueva_estruc/702825086213.pdf
- OIE. (2022). Modelo de cualificación y metodología de diseño de perfiles profesionales. Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura. Obtenido de <https://oei.int/wp-content/uploads/2022/09/20220825-modelo-y-metodologia-cnc-5.pdf>
- Pérez, G. R. (20 de Febrero de 2025). Cuatro de cada Diez trabajadores españoles deberá mejorar sus habilidades para tener UN empleo en 2030. Obtenido de El País: <https://elpais.com/economia/2025-02-20/cuatro-de-cada-diez-trabajadores-espanoles-debera-mejorar-sus-habilidades-para-tener-un-empleo-en-2030.html>
- Rama, C. (2009). El nacimiento de la acreditación internacional. Avaliação: Revista da Avaliação da Educação Superior (Campinas), 291-311. Obtenido de <http://www.scielo.br/pdf/aval/v14n2/a04v14n2>
- SPF. (2024). Exposición de Motivos: Paquete Económico 2025. (G. d. Puebla, Editor) Obtenido de Secretaría de Planeación y Finanzas del Estado de Puebla: <https://paqueteeconomico.puebla.gob.mx/2025/Exposiciondemotivos/>
- Tuduri, A. (17 de enero de 2025). Se dispara la demanda de expertos en data science. Obtenido de La Salle Campus Barcelona: <https://blogs.salleurl.edu/es/se-dispara-la-demanda-de-expertos-en-data-science>
- UNESCO. (2021). Remaining our futures together: A new social contract for education. Paris, Francia: United Nations Educational, Scientific and Cultural Organization. Obtenido de <https://unesdoc.unesco.org/ark:/48223/pf0000379707>
- UNESCO. (14 de Febrero de 2025). Educación superior. Obtenido de UNESCO: Building Peace through Education, Science and Culture, communication and information: <https://www.unesco.org/es/higher-education>

- Universidad Internacional de Andalucía. (2020). Estudio perfiles profesionales y habilidades asociadas. Andalucía, España. Obtenido de https://www.unia.es/images/observatorios/perfiles-profesionales/2020_Observatorio_Perfiles_completo.pdf
- UPlanner. (19 de Julio de 2022). ¿POR Qué la acreditación es importante en la educación superior? . Obtenido de uPlanner: <https://uplanner.com/es/acreditacion-educacion-superior/>
- Velasco, A., & Aguado, S. (2010). Modelos internacionales de certificación de competencias profesionales: una aplicación en la región de Madrid. Obtenido de https://www.aepro.com/files/congresos/2010madrid/ciip10_0483_0505.2793.pdf

Capítulo 10

Desarrollo de un sistema hipermedia adaptativo para potenciar capacidades de aprendizaje en matemáticas financieras de nivel superior

Eugenia Erica Vera Cervantes¹, Josefina Guerrero García², José Areli Carrera Román³

^{1,2} Benemérita Universidad Autónoma de Puebla,

³ Universidad Tecnológica de Tecamachalco

¹eugenia.vera@correo.buap.mx,

²josefina.guerrerogarcia@viep.com.mx,

³joseareli.carreraroman@viep.com.mx

Resumen. Las Matemáticas Financieras representan un componente esencial en la formación profesional, especialmente en disciplinas vinculadas con la administración, la contabilidad y la economía. No obstante, su carácter abstracto y su estrecha relación con el pensamiento lógico-matemático dificultan su comprensión para un número considerable de estudiantes. Por lo tanto, es indispensable desarrollar herramientas para reducir las carencias educativas. La meta es crear un Sistema Hipermedia Adaptativo (SHA), que toma en cuenta las particularidades individuales de los alumnos con el propósito de ajustar y ofrecer contenidos que fomenten el desarrollo de competencias en matemáticas financieras. En la parte metodológica de este trabajo se tiene la Fase1 de diagnóstico en la que se identificó el perfil académico a 66 estudiantes. En la Fase2 y 3 se diseñaron las estrategias de aprendizaje tomando en cuenta la propuesta de Bernad (2007) y la investigación-acción-desarrollo, se observaron y analizaron datos con la participación de 10 estudiantes de prácticas profesionales.

Palabras Clave: Matemáticas financieras, sistemas, adaptativo.

1 Introducción

Organizaciones internacionales como la Organización de las Naciones Unidas (ONU), han implementado programas para mejorar la calidad educativa ya que a través de la educación se pueden abordar y combatir problemas sociales, obstáculos culturales e incluso desafíos económicos y construir un futuro mejor para todos (UNESCO, 2023).

Tener conocimientos previos como métodos de aprendizaje y estilos de aprendizaje posibilita ajustar los recursos y las técnicas pedagógicas para atender requerimientos particulares. Dar los conocimientos personalizados ayuda al estudiante a mejorar sus aprendizajes, centrando, entonces, en el crecimiento integral de cada alumno, identificando sus puntos fuertes y áreas de mejora. Además, mejora la relación alumno-profesor, lo que robustece la relación de aprendizaje (Wilson et al., 2017).

Además, Ennouaman y Mahani (2017) hacen hincapié en que los materiales educativos deben estar disponibles para todos, pero también hay que asegurarse de que el aprendizaje sea el adecuado en cuanto a lugar, momento y método.

En las investigaciones de Carrera et al. (2023), Salcedo et al. (2016) y Cueli et al. (2016) se desarrollaron Sistemas Hipermedia Adaptativos (SHA), para mejorar el proceso de aprendizaje en matemáticas, y hablan sobre su la experiencia de adaptar los contenidos educativos al estudiante, identificando sus saberes anteriores, estilos de aprendizaje singulares y sus capacidades.

En la práctica educativa adaptar los materiales de aprendizaje se ha vuelto una práctica importante especialmente en el contexto actual, que demanda cambios imprescindibles en el proceso de enseñanza-aprendizaje.

En este escenario, los sistemas hipermedia adaptativos surgen como una respuesta de gran valor. Estos sistemas tienen la capacidad de ofrecer la flexibilidad necesaria para adaptar los contenidos educativos de forma individualizada, teniendo en cuenta las distintas capacidades y características de cada estudiante. Por lo tanto, se puede asegurar un camino claro para alcanzar los objetivos educativos establecidos (Vera et al., 2019).

El objetivo de esta investigación es proporcionar la adaptación y un aprendizaje más efectivo de las matemáticas financieras mediante el uso de un Sistema de Hipermedia Adaptativo. El trabajo mostrado se centrará en la elaboración de este Sistema Hipermedia.

En el apartado 1 se encuentra la introducción, enseguida en el apartado 2 se encuentra el estado de arte; en el apartado 3 el marco conceptual del estudio; en el apartado 4, la metodología establecida en la que se muestra la recopilación de los datos necesarios que guíen el desarrollo del SHA (diagnóstico); en el apartado 5, se presentan los resultados del diagnóstico antes de llevar a cabo el sistema propuesto. El desarrollo del SHA, en el apartado 6, enfatizando el beneficio que tendría un Sistema Adaptativo para el público propuesto. En el apartado 7, se discuten las conclusiones derivadas del estudio.

2 Estado del arte

En la actualidad la adaptación de contenidos se ha dado como respuesta a la heterogeneidad de los perfiles estudiantiles en áreas de matemáticas financieras. Este tipo de estrategias busca ajustar la secuencia y presentación de los materiales didácticos según el nivel de conocimientos previos y el estilo cognitivo del alumno, con el propósito de mejorar el entendimiento a procesos complejos como amortización, tasas de interés o valor presente neto.

En la literatura podemos encontrar investigaciones como las de Huang y Hew (2022) que aseguran que los entornos adaptativos en temas de matemáticas financieras favorecen el aprendizaje autorregulando y mejorando la motivación académica, como las de García-Peñalvo y Corell (2021) que hablan sobre la adaptación de recursos digitales en programas de finanzas promueve un aprendizaje más inclusivo y eficaz, al vincular teoría matemática con escenarios prácticos; las de Park y Jo (2023) han demostrado que los sistemas de aprendizaje personalizados en educación financiera reducen las brechas de rendimiento entre estudiantes con diferentes antecedentes académicos y fortalecen la toma de decisiones.

Entonces la evidencia reciente confirma que este enfoque posibilita el diseño de trayectorias formativas diferenciadas que favorecen la comprensión profunda de los conceptos, así como también facilitan la aplicación efectiva de los aprendizajes en contextos reales tanto de carácter económico como empresarial.

3 Marco teórico

La enseñanza de las Matemáticas Financieras demanda competencias lógico-matemáticas y posee un alto grado de abstracción, lo cual representa una dificultad recurrente para los estudiantes, esto hace que algunos avancen más que otros, de aquí la necesidad de implementar un Sistema Hipermedia Adaptativo (SHA).

Investigaciones recientes señalan que la falta de recursos personalizados limita la comprensión de conceptos como el valor del dinero en el tiempo, las tasas de interés y la evaluación de inversiones (Velásquez, 2021; UNESCO, 2023).

Este SHA integra recursos multimedia y ajusta contenidos en función de los perfiles académicos, los estilos de aprendizaje y las competencias de cada estudiante; de esta manera atiende la diversidad cognitiva, favorece la resolución de problemas financieros complejos, fomenta la autonomía e incrementa la motivación.

Estudios recientes enfatizan que los sistemas adaptativos aumentan la retención de conocimientos y promueven aprendizajes significativos (Wilson y Scott, 2017; Carrera, et al, 2023). De esta manera, el SHA se convierte en una herramienta pedagógica fundamental para mejorar el rendimiento académico y la preparación laboral de los estudiantes de finanzas, de esta manera, cubriendo las necesidades locales y las tendencias globales en innovación educativa.

3.1 Estudio de matemáticas financieras

Para Ramírez et al. (2009), las Matemáticas Financieras representan un grupo de conceptos y métodos de análisis, efectivos para la comparación y valoración económica de opciones, donde su propósito es tomar la mejor decisión desde una perspectiva económica.

Las matemáticas financieras se han convertido en una cuestión fundamental para la toma de decisiones, cuando se destinan recursos a proyectos o inversiones, así como facilitar la determinación del valor del dinero a lo largo del tiempo (Velásquez, 2021).

Esta investigación se fundamentó en el currículo 2019 de la asignatura de Matemáticas Financieras de Finanzas de una universidad pública, dividiendo los contenidos en cuatro secciones, cada una centrada en aclarar de forma clara conceptos

fundamentales, solucionar ejercicios prácticos, instruir en el manejo y despeje de fórmulas y potenciar el aprendizaje mediante actividades diseñadas de forma pedagógica que involucren soluciones de problemas reales, la vinculación con escenarios de la vida real y la utilización de tecnología y herramientas digitales para representar y analizar conceptos financieros.

3.2 Sistema hipermedia adaptativo

Los Sistemas Hipermedia Adaptativos basados en inteligencia artificial son una herramienta innovadora y poderosa que permiten un aprendizaje más eficiente en educación superior, personalizando contenidos según el rendimiento y la motivación del estudiante (Li et al., 2025). Torres y Jiménez (2025) evidencian que la integración de sistemas hipermedia en matemáticas financieras mejora la comprensión conceptual y la toma de decisiones en escenarios simulados. Mientras que Kumar y Singh (2025) demuestran que estas herramientas fortalecen las competencias digitales y reducen la ansiedad matemática al ofrecer rutas de aprendizaje dinámicas y diferenciadas.

Mediante un diseño óptimo, es posible proponer contenido hipermedia, como videos, imágenes, texto y enlaces, que se presenten de diversas formas según las necesidades del aprendiz, por lo que es indispensable un diagnóstico previo (Vera, et al, 2019).

La capacidad de adaptación en un sistema hipermedia adaptativo significa que el programa modificará su contenido, estructura y presentación, basándose en las particularidades del usuario. Se consigue esto recolectando y evaluando datos acerca del usuario, tales como su perfil, historial de interacciones, preferencias, grado de conocimiento, contexto situacional, entre otros aspectos (Peña, 2019).

Un sistema hipermedia adaptativo puede estar formado por diversos componentes que colaboran para ofrecer una experiencia personalizada al usuario, compuesto por tres modelos (dominio, usuario y adaptación). El modelo de usuario es la fuente de información que constituye una representación organizada de los datos recolectados sobre el usuario, e incluye diferentes características, que dependen de los alcances propuestos para la construcción del sistema (Wilson y Scott, 2017).

El modelo de adaptación constituye el elemento esencial del sistema hipermedia adaptativo, Mediante el uso de métodos de inteligencia artificial, tales como la minería de información y el aprendizaje automático, el motor de adaptación examina el modelo de usuario y decide cómo ajustar el contenido y la presentación. Es posible elegir el contenido pertinente, modificar la secuencia de los enlaces y personalizar la interfaz de usuario.

3.3 Perfil académico

El perfil académico del estudiante (entendido como el conjunto de conocimientos, habilidades y características cognitivas que definen su desempeño esperado en un contexto formativo) es un factor clave para diseñar estrategias de intervención efectivas. Es necesario una perspectiva integradora del perfil académico del estudiante, esto se puede obtener a través de la utilización de diversos instrumentos por una vía que le facilite la aplicación (Sarmiento 2007).

En nuestra investigación, se toma este perfil académico como fuente de información para el SHA para el aprendizaje autónomo de matemáticas financieras. Este sistema requiere estar adecuadamente informado sobre las características académicas de los alumnos que se registren. Se eligieron cinco variables fundamentales con sus correspondientes dimensiones, entre las que se incluyen: estilo de pensamiento, vía de preferencia, método de aprendizaje, grado de lectura y nivel de conocimientos. Se estimó cada una de estas variables a partir de la utilización en línea de instrumentos adaptados para su empleo en este medio. Se utilizaron herramientas validadas en sus versiones iniciales y, principalmente, de conocimiento y uso habitual por investigadores y profesores.

Para la evaluación de los estilos de aprendizaje se utilizó el cuestionario HONEY-ALONSO, el cual clasifica a los estudiantes en estilos activos, reflexivos, teóricos y pragmáticos (Honey y Alonso, 1997). Las competencias en lectocomprensión se evaluaron utilizando una versión modificada del Test de Habilidades Lectocomprensivas Básicas de Riquelme, Constenla y Cruces (1996), y el grado de conocimientos a través de un examen protocolo cuya estructura fue sugerida por Bernad (2007). En cuanto al Razonamiento Científico (tipo de pensador), se aplicó una versión modificada del Examen de Clase de Razonamiento Científico de Lawson (1978). Esta prueba se fundamenta en la teoría del desarrollo de la inteligencia de Piaget y facilita la identificación de los alumnos que la solucionan en uno de tres niveles posibles: el pensador concreto, el pensador transitorio y el pensador formal.

4 Metodología

Esta investigación se llevó a cabo usando la Investigación-Acción-Desarrollo (IAD), siguiendo el modelo en espiral de Kemmis, McTaggart y Nixon (2014). La parte importante de este método es como combina la investigación sistemática con la acción pedagógica, logrando una transformación, integrando la planificación, la puesta en marcha de estrategias, la observación reflexiva y la mejora constante en situaciones reales de enseñanza y aprendizaje.

La IAD es una estrategia metodológica de carácter cualitativo y participativo que busca generar conocimiento a través de la transformación de la práctica educativa (Kemmis, et al., 2014).

En este enfoque los procesos de enseñanza-aprendizaje requieren ser abordados desde el interior mismo de la práctica, con la participación activa de los sujetos involucrados (Kemmis, et al., 2014).

Se compone por cuatro etapas recurrentes: planificación, acción, observación y reflexión crítica. Cada una de estas fases constituye un ciclo que puede repetirse sucesivamente en función de los hallazgos obtenidos, lo que permite una mejora continua del objeto de estudio en contextos reales. La planificación implica identificar un problema o necesidad contextualizada; la acción corresponde a la intervención diseñada para atender dicha problemática; la observación permite recolectar datos sobre la implementación; y la reflexión crítica, promueve el análisis colectivo de los resultados con miras a reformular nuevas estrategias (ver figura 1).



Figura 1. Esquema de Investigación Acción Desarrollo basado en el modelo de Kemmis, McTaggart y Nixon (2014).

En nuestro caso, se estructuraron tres fases ejecutadas y una fase proyectada, como se detalla a continuación y los resultados del diagnóstico se muestran en la sección 5:

Fase 1. Planificación

En esta fase, buscamos conocer el perfil académico actual del estudiante, descrito en el marco teórico y saber qué tipo de actividades asignar. En esta fase participaron 66 estudiantes de Matemáticas Financieras de una universidad pública. La información permitió formular una base diagnóstica con relación a los contenidos de mayor dificultad y a las competencias a desarrollar con mayor prioridad, canal de preferencia, estilo de aprendizaje, tipo de pensador y nivel de lectura.

Fase 2. Acción

En esta etapa se diseñaron e implementación las actividades de aprendizaje del Sistema Hipermedia Adaptativo (SHA), las cuales se elaboraron a partir de los resultados del diagnóstico inicial, y se fundamentaron en dos enfoques metodológicos: la IAD, que guio la construcción iterativa de los recursos, y la Escala de Estrategias de Aprendizaje Contextualizado (ESEAC) de Bernad (2007), que permitió enriquecer el contenido didáctico.

Las actividades diseñadas se centraron en promover la comprensión conceptual, el uso de fórmulas y la resolución de problemas aplicados en contextos reales. La idea es fomentar la autonomía, la autorregulación y el uso de materiales personalizados.

Fase 3. Observación

Durante esta etapa se observó y documentó la experiencia de uso del SHA por parte de un grupo focal de 10 estudiantes de finanzas que cursaron la materia de Matemáticas Financieras, mediante notas de profesor y vídeos. Realizando una reflexión sobre las actividades aplicadas. Esta observación permitió identificar áreas de mejora en el diseño instruccional, en la organización del contenido, en la accesibilidad y en la usabilidad de la interfaz.

Fase 4. Reflexión crítica (propuesta futura)

La cuarta fase del ciclo de IAD, correspondiente a la reflexión crítica, está prevista como parte del siguiente ciclo metodológico y representa una etapa proyectada del estudio. Esta fase contemplará la evaluación formal de la efectividad del SHA como herramienta de apoyo en el aprendizaje de las Matemáticas Financieras, mediante la comparación entre grupos de intervención y control, así como el análisis del impacto en el desempeño académico de los estudiantes.

5 Resultados del diagnóstico previo al desarrollo del SHA

A través de la Fase 1, donde participaron 66 alumnos inscritos en la materia de Matemáticas Financieras del Plan de estudios perteneciente a una institución pública de nivel superior, se realizó una Prueba Protocolo sobre el estado actual de los estudiantes, el cual se compone de 25 reactivos que proporcionaron información sobre los temas de mayor dificultad, las competencias a desarrollar. En la tabla 1 se describen los resultados obtenidos de la prueba protocolo y en la tabla 2 los resultados de los instrumentos aplicados descritos en el marco teórico. De esta forma, este diagnóstico situacional constituyó la base para poder establecer actividades de aprendizaje adecuadas y adaptadas a cada uno de los perfiles contextuales.

Tabla 1. Resultados de la prueba protocolo.

Dominio	Rango	Media	Interpretación
General del tema (D1)	2.5-3.4	3	El nivel de conocimiento de los estudiantes en la solución de problemas es regular.
De lenguaje (D2)	3.5 – 4.4	3.9	Los estudiantes se expresan de manera clara y gestionan de manera coherente las diferentes manifestaciones del lenguaje icónico, verbal y escrito, logrando un resultado positivo (Bueno).
Calidad de Razonamiento (D3)	2.5-3.4	2.4	Los estudiantes incurrir en fallos significativos al hacer inferencias y argumentaciones, y surgen fallos en la implementación de la respuesta, resultando en un resultado deficiente (Insuficiente).
Errores y su naturaleza (D4)	2.5– 3.4	3	Los estudiantes presentan ciertas dudas e inconsistencias al realizar la respuesta, aunque en ciertos aspectos resulta claramente lógico. En donde el estudiante no está reconociendo sus fallos y necesita perfeccionar su observación y concentración (Insuficiente).
Nivel de Abstracción (D5)	1.5-2.4	2.3	Esto nos señala que los estudiantes no son capaces de construir una representación mental de la realidad, por lo que incurrir en errores significativos al realizar tales representaciones, siendo en general suficiente.
Conciencia Cognitiva (D6)	2.5-3.4	3	Esto nos señala que los estudiantes no son capaces de construir una representación mental de la realidad, por lo que incurrir en errores significativos al realizar tales representaciones, siendo en general suficiente (Regular).

Nivel de motivación y ansiedad (D7)	1.5-2.4	2.3	Los estudiantes presentan ciertas incertidumbres e inconsistencias al gestionar sus emociones para llevar a cabo tareas o actividades, a pesar de que en ciertos aspectos se autoregulan y a su vez necesitan automotivación.
-------------------------------------	---------	-----	---

Tabla 2. Resultados de los instrumentos.

Instrumentos	Resultados
Tipo de pensador	Concreto 30.3%. En transición %65.6. Formal 4.1%
Canal preferente de aprendizaje	Visual 31.8%. Auditivo 21.2%. Kinestésico 47%
Estilo de aprendizaje	Activos 34.8%. Reflexivos 33.3%. Teóricos 15.2%. Pragmáticos 16.7%
Lectura y comprensión	Baja 30%. Media 60%. Alta 10%

En la Fase 2 se elaboraron las actividades de aprendizaje tomando como referencia el diagnóstico obtenido en la Fase 1, adicionalmente los resultados de las acciones de investigación-acción-desarrollo, participaron 10 estudiantes de finanzas, se aplicaron las actividades diseñadas de los temas que representaban mayor dificultad, arrojando información relevante sobre áreas de mejora en el tipo de materiales propuestos. En la Fase 3 atendiendo a esa información se hicieron ajustes a las actividades y posteriormente se implementaron en el SHA. En el desarrollo de estas actividades se consideró: *tipo de aprendizaje, tipo de pensador, canal preferente de aprendizaje, estilos de aprendizaje y competencias a desarrollar* (ver tabla 3). Este diseño integral asegura que las actividades respondan a las necesidades individuales de los estudiantes, optimizando su desempeño en Matemáticas Financieras.

Tabla 3. Actividades de aprendizaje y sus atributos.

Unidad 1		Conceptos básicos													Perfil del estudiante al que va dirigida la actividad			
Actividad			Competencia		Estrategia		Tipo de aprendizaje (Quesada,1996)		Proceso cognitivo (Bloom, 2004)									
Tema	Subtema	Actividad	Genérica	Específica	Atributo	Estrategia	Antecedentes	Indispensable	Esencial	Conocimiento	Comprensión	Aplicación	Análisis	Síntesis	Evaluación	Tipo de pensador	Canal preferente de aprendizaje	Estilos de aprendizaje
1.	1.6.1	22	4	7	28	7	X					X				C	K	A
1.	1	2	4	7	28	10	X					X				C	K	A
1.	1.1	3	4	7	28	10		X		X						T	V	T
1.	1.1	4	4	7	28	10						X				C	V	P
1.	1.1	5	4	7	28	10	X					X				C	K	A
1.	1.1	6	4	7	28	10			X			X				C	K	T

El **tema** indica la unidad de aprendizaje, el **subtema** indica el contenido temático, la **competencia** indica la competencia genérica y específica a desarrollar del plan de estudios de la materia de Matemáticas Financieras. La **estrategia**, son estrategias de aprendizaje contextualizado de Bernad (2007), las cuales trabajan con la parte: cognitiva, estratégica y contextualizada, cada una de ellas representa un producto o meta parcial y diferente claramente objetivable dentro de cada dimensión propuesta. Por ejemplo, se tiene la Estrategia 3ª que permite codificar los contenidos que procesa,

otra para relacionar los datos representados mentalmente en una fase anterior de procesamiento (Estrategia 4^a). Y los valores de los atributos: *tipo de aprendizaje, proceso cognitivo, tipo de pensador, canal preferente, estilos de aprendizaje, lectura y comprensión*, que se encuentran en la tabla 4.

Tabla 4. Atributos de las actividades de aprendizaje

Tipo de Aprendizaje	Proceso intelectual	Tipo de pensador	Canal preferente aprendizaje	Estilos de aprendizaje	Lectura y comprensión
A=Antecedente I=Indispensable E=Esencial	1. Conocimiento 2. Comprensión 3.- Aplicación 4.- Análisis 5.- Síntesis 6.- Evaluación	C= Concreto T= Transición F= Formación	A= Auditivo V= Visual K= Kinestésico	A=Activo R=Reflexivo T=Teórico P=Pragmático	B= Baja M= Media A= Alta

El SHA fue diseñado con base en los principios del enfoque constructivista, lo que implicó la creación de diversas modalidades de clase orientadas a promover una participación del estudiante en la construcción de su conocimiento. La estructura de cada clase ha sido cuidadosamente diseñada para facilitar la articulación entre los saberes previos del alumno y los nuevos contenidos, en un entorno de aprendizaje contextualizado y cercano a la realidad.

La Fase 4, se llevará a cabo con estudiantes que presentan dificultades en la materia, este reporte se presentará en una publicación posterior.

6 Diseño del SHA

Basándonos en el análisis de las Fases 1, 2 y 3 de intervención, logramos desarrollar las actividades de aprendizaje con las características que muestra la tabla 4.

Para el estudio y desarrollo del sistema, se realizaron los casos de uso mediante Lenguaje de Modelado Unificado (UML) conforme a Ciccozzi, et al.(2019). Este es un lenguaje de modelado estándar empleado en la ingeniería de software para visualizar, especificar, construir y registrar sistemas de software. Este modelado ofrece un conjunto de anotaciones gráficas y recursos para ilustrar visualmente distintos elementos de un sistema, tales como su estructura, conducta, interacciones y vínculos entre sus partes. Un sistema hipermedia adaptativo (SHA) engloba a diversos participantes fundamentales que se relacionan entre ellos para ofrecer una experiencia de aprendizaje a medida, los cuales son: *el estudiante, el docente, el gestor y la base de datos*.

Base de datos

El SHA incluye una base de datos relacional, la cual ha sido meticulosamente elaborada mediante un modelo conceptual. Para alcanzar este objetivo, primero se recolectaron todos los requerimientos para el funcionamiento adecuado del sistema. Luego, se elaboró un esquema exhaustivo de la base de datos que abarca descripciones de las entidades, relaciones y limitaciones, que se muestra empleando conceptos gráficos y textuales característicos del modelo de datos relacionales. Se perfeccionó hasta crear

un esquema lógico. En el producto se edifica y administra la información de forma eficaz y segura. En esta base de datos se almacenan todas las actividades con los atributos explicados de la tabla 3, así como se registra el avance del alumno y la próxima vez que el inicie sesión no tendrá que responder preguntas de temas pasados, si no que responderá la pregunta correspondiente al tema y subtema que se encuentra cursando.

Descripción de contenido

Para la creación del sistema hipermedia adaptable, se crearon varias *clases* con el objetivo de cubrir las demandas de cada tipo de estudiante. Estas clases están accesibles para su estudio, facilitando a todos los estudiantes obtener los conocimientos requeridos en un ambiente envoltorio próximo a la realidad. Así, los estudiantes pueden crear vínculos entre los nuevos saberes y los ya adquiridos, además de enfrentar problemas futuros.

Al ingresar al SHA, el estudiante tiene la posibilidad de examinar y elegir su asignatura disponible, en este caso: *matemáticas financieras*. Al acceder al curso, la primera pantalla que se muestra es la de inicio de sesión, donde el estudiante puede ingresar utilizando su cuenta.

El SHA ha sido elaborado conforme a la base de los principios del constructivismo, lo que ha implicado la elaboración de modalidades de clase orientadas a una participación del alumno para la construcción de su propio conocimiento. La configuración de cada una de las clases ha sido diseñada para facilitar la construcción de la articulación entre los saberes previos del alumno y los saberes nuevos, todo ello en un ambiente de aprendizaje contextualizado y próximo a la realidad.

Cada una de las clases contenidas en el SHA está intercalada por etapas que permiten ir interactuando con los contenidos. Así el alumno parte de observar la clase que corresponde al tema, que está disponible en escritos, en materiales visuales o audiovisuales. Una vez el alumno considera la información, puede acceder a una actividad orientada a practicar lo que ha aprendido. Esta actividad debe ser resuelta de forma autónoma, y posteriormente el sistema genera una retroalimentación o calificación inmediata, fortaleciendo así el aprendizaje autorregulado.

Los tipos de clases integradas en el SHA incluyen: **Debate:** Fomenta el pensamiento crítico mediante la confrontación de ideas. **Desarrollo del tema:** Presenta la explicación estructurada de los contenidos. **Conceptos clave:** Ofrece definiciones y elementos fundamentales del tema. **Video del tema:** Recurso audiovisual que refuerza la comprensión mediante ejemplos prácticos.

Estos materiales forman parte de un conjunto de Recursos Educativos Abiertos (REA), seleccionados y organizados para garantizar la enseñanza de los conceptos esenciales en Matemáticas Financieras (ver Figura 2).



Figura 2. Clases por debate y video del tema.

Para lograr que los materiales sean un verdadero apoyo didáctico de calidad, se presentan al usuario diversas actividades de aprendizaje sencillas que involucran los procesos cognitivos siguientes: Interpretación de información, aplicar la idea principal y comprender la idea esencial de cada tema. Se diseñaron cinco tipos de actividades que se describen a continuación: *Booleana*, *Opción Múltiple*, *Completar Texto*, *Relacionar* y *Arrastrar* (ver figura 3).

En cada actividad es posible que el alumno se encuentre con un poco de indecisión al momento de responder a su actividad, en este caso todas las actividades cuentan con la opción de regresar a la clase que observaron antes de dicha actividad, funcionando como una ayuda para resolver dudas, sin otorgar directamente la respuesta.

Una vez contestada la actividad por parte del alumno se procede a la evaluación. En caso de que la respuesta haya sido incorrecta el alumno recibirá en pantalla un tip de ayuda con una idea clave que le servirá para recordar los conocimientos recibidos en la clase y vincularlos, esta vez de manera correcta.



Figura 3. Actividad Verdadero o Falso y actividad “Completar frase”.

7 Conclusiones

Se desarrolló un SHA, en el que se ponen en operación actividades de aprendizaje de corte autónomo, con un enfoque constructivista centrado en el perfil académico que mostraron los estudiantes que se encuentran cursando la carrera de Finanzas en una universidad pública, derivadas del análisis de los factores de rendimiento académico

que señala la literatura especializada y un diseño didáctico de éstas basado en la investigación-acción-desarrollo y la escala ESEAC, del cual se espera ayude a fortalecer el proceso de enseñanza-aprendizaje en la asignatura de Matemáticas Financieras.

Se sugiere la puesta en marcha del SHA con las generaciones venideras de alumnos y profesores de la carrera de Finanzas. Esta aplicación facilitará la comprobación de su eficacia en situaciones reales, la recolección de comentarios útiles y la realización de las modificaciones requeridas para mejorar su funcionalidad. Dentro de las mejoras proyectadas se incluye la inclusión de escenarios más fieles, interfaces más fácilmente comprensibles y sistemas de personalización sofisticados. Este trabajo constituye un avance inicial en la creación de sistemas educativos digitales flexibles para la capacitación en Matemáticas Financieras en la carrera de Finanzas. A mediano plazo, se implementará esta propuesta a otras materias del currículo, con la finalidad de construir un ecosistema de objetos de aprendizaje que ayude de forma sistemática a mejorar el desempeño escolar y a potenciar las habilidades profesionales de los alumnos.

Referencias

- Bernad, J.A. (2007) Modelo Cognitivo de Evaluación Educativa, Escala de Estrategias de aprendizaje Contextualizado (ESEAC), Narcea, S.A. de Ediciones Madrid.
- Carrera, J.A. Guerrero, J. y Vera, E.E (2023). Design of an adaptive hypermedia system to strengthen algebra learning skills of high school students. *Journal of Teaching and Educational Research*. Vol. 9(24), pp.13-24.
- Ciccozzi, F., Malavolta, I., y Selic, B. (2019). Execution of UML models: A systematic review of research and practice, *Software & Systems Modeling*, 18, 2313-2360. <https://doi.org/10.1007/s10270-018-0675-4>
- Cueli, M., González-Castro, P., Krawec, J., Carlos Núñez, J., y Antonio González-Pienda, J. (2016). Hipatia: A hypermedia learning environment in mathematics. *Anales de Psicología*, 32(1), 98–105. <https://www.redalyc.org/articulo.oa?id=16743391011>
- Ennouamani, S. y Mahani, Z. (2017). An overview of adaptive e-learning systems. Conference: Eighth International Conference on Intelligent Computing and Information Systems (ICICIS).
- García-Peñalvo, F. J., y Corell, A. (2021). La adaptación de contenidos digitales en la educación superior: retos y oportunidades. *Education in the Knowledge Society*, 22(1), 1–12. <https://doi.org/10.14201/eks.23495>
- Honey, P., y Alonso, C. M. (1997). *Los estilos de aprendizaje: Procedimientos de diagnóstico y mejora*. Bilbao: Ediciones Mensajero.
- Huang, R. H., y Hew, K. F. (2022). Implementing adaptive learning systems in finance and mathematics education: Effects on self-regulation and achievement. *Interactive Learning Environments*, 30(6), 1021–1038. <https://doi.org/10.1080/10494820.2020.1815223>
- Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: Doing critical participatory action research*. Springer. <https://doi.org/10.1007/978-981-4560-67-2>
- Kumar, R., y Singh, P. (2025). Adaptive learning pathways to reduce math anxiety in financial mathematics. *Journal of Educational Technology Research*, 48(2), 145–168.
- Lawson, A. E. (1978). Development and validation of the classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15(1), 11–24. <https://doi.org/10.1002/tea.3660150103>

- Li, X., Chen, Y., y Zhao, L. (2025). Adaptive hypermedia systems for higher education: Enhancing learning efficiency with artificial intelligence. *Computers & Education*, 210, 104882. <https://doi.org/10.1016/j.compedu.2025.104882>
- Park, Y., y Jo, I. H. (2023). Personalized learning in financial mathematics: Reducing achievement gaps through adaptive pathways. *Journal of Computer Assisted Learning*, 39(4), 1112–1127. <https://doi.org/10.1111/jcal.12789>
- Peña, K. A. (2019). “Sistema de gestión de conocimiento para redes académicas apoyado en sistemas de hipermedia adaptativos”, Recuperado de: <http://hdl.handle.net/11349/15090>.
- Ramírez, C., García, M., Pantoja, C. y Zambrano A. (2009). *Fundamentos de las Matemáticas Financieras*. Editorial Universidad Libre Sede Cartagena.
- Salcedo, P., López, O., y del Valle, M. (2016). LEXMATH: Un hipermedio adaptativo para el aumento del léxico en matemática. *RECHIEM. Revista Chilena de Educación Matemática*, 10(1), 106-112.
- Sarmiento, M. (2007). “La enseñanza de las matemáticas y las Ntic”, Una estrategia de formación permanente, (Tesis la Universitat Rovira i Virgili, España). Repositorio T.1625-2007.
- Torres, A., y Jiménez, P. (2025). Aplicación de sistemas hipermedia adaptativos en matemáticas financieras: un enfoque experimental en educación universitaria. *Revista Iberoamericana de Tecnología Educativa*, 20(1), 55–72.
- UNESCO (2023). “El trabajo de la UNESCO en el ámbito de la educación”, UNESCO, Recuperado de: <https://www.unesco.org/es/education/action>
- Velásquez, V. J. (2021). La base conceptual de las matemáticas financieras para la toma de decisiones, *Investigación y Negocios*, vol. 14(24), pp. 135-148. http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2521-27372021000200135
- Vera, E., Navarro, Y., y Guerrero, J. (2019). Sistema hipermedia adaptativo para el apoyo del aprendizaje autónomo del idioma inglés, *Research in Computing Science*, vol. 148(3), 91-104. <https://doi.org/10.13053/rscs-148-3-8>
- Wilson, C., y Scott, B. (2017). Adaptive systems in education: a review and conceptual unification. *The International Journal of Information and Learning Technology*. <https://doi.org/10.1108/IJILT-09-2016-0040>

Índice de autores

Berenice Rodríguez Pedroza	Juan Carlos Conde Ramírez
Carlos Armando Ríos Acevedo	Luis Enrique Morales Aguilar
Carlos Leopoldo Carreón Díaz de León	María Teresa Torrijos Muñoz
Daniel Marcelo González Arriaga	Mauro Alberto López Muñoz
Eugenia Erica Vera Cervantes	Mireya Tovar Vidal
Héctor David Ramírez Hernández	Nelva Betzabel Espinoza Hernández
Héctor Gabriel Ramírez Espinoza	Regina Monserrat Avelino Farfán
Jaime Alejandro Romero Sierra	Roberto Contreras Juárez
José Areli Carrera Román	
Josefina Guerrero García	

Compiladores

Mireya Tovar Vidal
María Teresa Torrijos Muñoz
Carlos Leopoldo Carreón Díaz de León
Daniel Marcelo González Arriaga
Juan Carlos Conde Ramírez
Jaime Alejandro Romero Sierra

Revisores

Ana Laura Lezama Sánchez	José Luis García Cue
Carlos García Franchini	José Luis Zechinelli Martini
Carlos Leopoldo Carreón Díaz de León	José Víctor Flores Flores
Claudia Zepeda Cortés	Juan Carlos Conde Ramírez
Daniel Marcelo González Arriaga	Julio Hernandez Torres
Erick Barrios González	María Adelina Cruz
Etelvina Archundia Sierra	María Patricia Torrijos Muñoz
Freddy Palma Mancilla	María Teresa Torrijos Muñoz
Gerardo Arizmendi Echegaray	Martha Alvarado Arellano
Héctor Saib Marabillo Gómez	Miguel Ocaña Bribiesca
Héctor David Ramírez Hernández	Milagros Zeballos Rebaza
Hilda Castillo Zacatelco	Mireya Tovar Vidal
Hugo Villanueva Méndez	Nelva Betzabel Espinoza Hernández
Jaime Alejandro Romero Siera	Reyna Carolina Medina Ramírez
José Alejandro Reyes Ortiz	Ruben Blancas Rivera
José Laguna Cortés	Zobeida Jezabel Guzmán Zavaleta
José Luis Carballido Carranza	

Editores

Mireya Tovar Vidal
María Teresa Torrijos Muñoz
Carlos Leopoldo Carreón Díaz de León
Daniel Marcelo González Arriaga
Juan Carlos Conde Ramírez
Jaime Alejandro Romero Sierra

Trascendencia de la didáctica de la Ciencia de Datos en el proceso de
enseñanza-aprendizaje a nivel superior
Editores de la publicación:
Mireya Tovar Vidal
María Teresa Torrijos Muñoz
Carlos Leopoldo Carreón Díaz de León
Daniel Marcelo González Arriaga
Juan Carlos Conde Ramírez
Jaime Alejandro Romero Sierra
A partir de diciembre de 2025
está disposición en PDF en la página
de la Facultad de Ciencias de la Computación
de la Benemérita Universidad Autónoma de Puebla (BUAP)
<https://icds.cs.buap.mx/static/LIBRO2TrascendenciaCD.pdf>
Peso del archivo: 6.00 MB

