



2022

Estrategias Didácticas **INNOVADORAS**

en Educación Superior

.....
Aprendiendo Machine Learning en la UdeA

Título de la estrategia

Estrategia didáctica para la enseñanza del Machine Learning con Python aplicada a estudiantes de los programas de Biología, Química, Estadística y Matemáticas

Unidad académica

Facultad de Ciencias Exactas y Naturales
(Biología, Estadística, Química y Matemáticas)

Autores

Investigador principal

Marco Julio Cañas Campillo

Licenciado en Matemáticas y Física, magíster en Ciencias Matemáticas, docente ocasional en la Seccional Bajo Cauca e investigador del Grupo de Investigación Interdisciplinaria Bajo Cauca y sur de Córdoba (GIBACC).

Coinvestigadores

Luz Mariela López Nohava

Bibliotecóloga, magíster en Educación, coordinadora de la Biblioteca Bajo Cauca, investigadora del Grupo de Investigación Interdisciplinaria Bajo Cauca y sur de Córdoba (GIBACC), y coordinadora del Semillero Interdisciplinario de Investigación del Bajo Cauca y sur de Córdoba (SIGIBACC) adscrito al GIBACC.

Autores

Agradecimientos especiales a **Yuber Hernany Tapias Arboleda**, licenciado en Matemáticas Básicas, magíster en Matemáticas y docente de cátedra, por el apoyo y los aportes a la estrategia didáctica innovadora.

Estudiante de apoyo

Janis Zúñiga Ortega

Estudiante de sexto nivel de Biología e integrante del Semillero Interdisciplinario de Investigación del Bajo Cauca y sur de Córdoba (SIGIBACC).

Créditos

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.



Índice



Glosario	6
Resumen	8
1. Contexto de la estrategia	9
2. Metodología de aprendizaje	13
3. Rol del docente y del estudiante	15
4. Factor diferenciador	16
5. Evidencias	17
6. Plan de evaluación	18
7. Factores críticos	21
Referencias bibliográficas	22

Haz clic en los números para ir a la página

Glosario

GitHub de Microsoft: plataforma de desarrollo colaborativo en línea que permite alojar y controlar versiones de proyectos de software. Es especialmente popular entre los desarrolladores de software y otros profesionales de tecnología, y se utiliza ampliamente para el almacenamiento y la gestión de código fuente y proyectos de software.

Machine Learning: campo de estudio que le otorga a las computadoras la capacidad de aprender sin ser programadas explícitamente. Se dice que un programa de computadora aprende de la experiencia E con respecto a alguna tarea T y alguna medida de desempeño P , si su desempeño en T , medido por P , mejora con la experiencia E (Mitchell et al., 1997).

Python: lenguaje de programación de propósito general y alto nivel. Su filosofía de diseño enfatiza la legibilidad del código con el uso de sangría significativa. Este lenguaje facilita la gestión y manipulación de datos, y la construcción de modelos predictivos de Machine Learning y Deep Learning.

Representación geométrica para datos: son formas geométricas de representar los datos y los modelos predictivos que aprenden de ellos. Algunas de las representaciones más comunes para datos incluyen:

- **Diagramas de barras:** útiles para comparar valores absolutos o relativos de distintas categorías.
- **Gráficos de línea:** adecuados para visualizar los cambios en una variable a lo largo del tiempo.
- **Gráficos de dispersión:** útiles para mostrar la relación entre dos variables.
- **Gráficos de torta:** útiles para mostrar la relación de partes de un todo.
- **Diagramas de árbol:** útiles para mostrar relaciones jerárquicas entre diferentes elementos.

Glosario

Antes de elegir una representación gráfica para tus datos es importante considerar el tipo de información que quieres mostrar y el público al que va dirigida. Algunos gráficos pueden ser más fáciles de interpretar que otros dependiendo del contexto y del conocimiento previo del público.

SciKit Learn: biblioteca de aprendizaje automático de software libre para el lenguaje de programación Python. Cuenta con varios algoritmos de clasificación, regresión y agrupamiento que incluyen máquinas de vectores de soporte, bosques aleatorios, aumento de gradiente, k-means y DBSCAN, y está diseñado para interactuar con las bibliotecas numéricas y científicas de Python NumPy y SciPy.



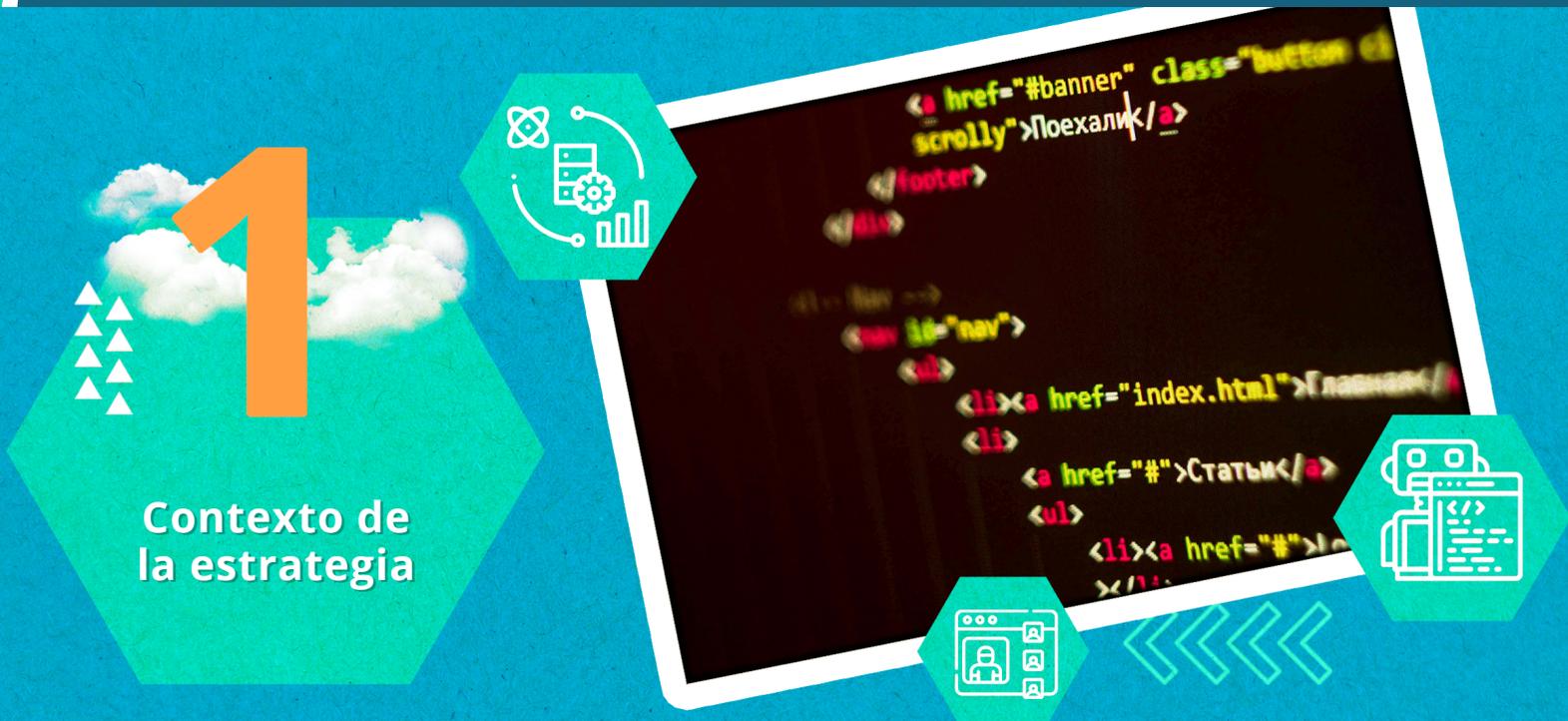
Resumen

Esta propuesta recoge la experiencia didáctica para la enseñanza de la ciencia de datos a través de dos cursos, *Machine Learning* e *Introducción al Machine Learning*, aplicada a estudiantes de noveno y décimo semestre de la Facultad de Ciencias Exactas y Naturales de la Universidad de Antioquia. Se presenta como una estrategia innovadora en la educación superior porque, ante la nueva realidad de educación virtual que potencia el uso de los recursos digitales, se estableció como un proceso basado en la constitución de un sistema articulado de recursos de enseñanza, apropiación del conocimiento y evaluación que facilitó la interacción en el aprendizaje y el desarrollo de habilidades, e involucró plataformas como GitHub de Microsoft, Google Colab, YouTube, Kaggle de Google y Jupyter, y Spyder de Anaconda. En consecuencia, se considera una propuesta exitosa que se podría replicar transversalmente en varios programas regionalizados de la Universidad.



El lenguaje de programación Python está orientado a la ciencia de datos. Las librerías Numpy, Pandas, Matplotlib y SciKit Learn de Python se consideran herramientas de manipulación, procesamiento, visualización y modelamiento.

Nota: Te invitamos a acompañar la lectura de esta cartilla con el [canal de YouTube DiMathData](#) y los [cuadernos de GitHub](#), donde encuentras todos los contenidos utilizados en el desarrollo de la estrategia didáctica.



En el marco de la cuarta revolución industrial, las universidades incursionan en la formación curricular de los estudiantes hacia la apropiación de conocimientos y el desarrollo de habilidades en la ciencia de datos; estos últimos son necesarios para la toma de decisiones acertadas en el mundo científico, industrial y empresarial. Dice Schwab que en esta revolución millones de los empleos nuevos serán para quienes tengan la formación y las competencias acordes a estas nuevas tendencias, por lo que el mayor desafío es “formar la fuerza laboral del futuro y, al mismo tiempo, ayudar a los trabajadores de hoy a hacer la transición a esta nueva economía” (2016, p. 6).

Asimismo, la Universidad de Antioquia (2017; 2021), plantea la innovación como un pilar fundamental que permite la asimilación y la adaptación a las tendencias globales de la ciencia y la tecnología para resolver problemáticas en clave de desarrollo territorial. En esa medida, los docentes deben formar a los estudiantes hacia la capacidad de diversificar la estructura productiva en el ámbito económico para responder a los retos de la transformación digital y generar nuevos conocimientos que demanda el país.

De acuerdo con lo anterior, la Universidad de Antioquia oferta asignaturas que apalancan los saberes de la ciencia de datos; no obstante, es a partir de la contingencia por el Covid-19 que el uso de las tecnologías para la modalidad de educación virtual se impone como una

opción ante la problemática del encierro obligatorio, pero, también, como una oportunidad de innovación didáctica que obligó al docente a reinventarse y a diseñar estrategias en la formación de los estudiantes para alcanzar objetivos concretos en el desarrollo de habilidades de pensamiento y competencias computacionales, las cuales facilitarán, de forma creativa, la apropiación de los conocimientos, el análisis de la información y la automatización de los datos (Wu, Li, Chandru y Daniel, 2022; Salas, Ramírez, Eslava, et al., 2022).

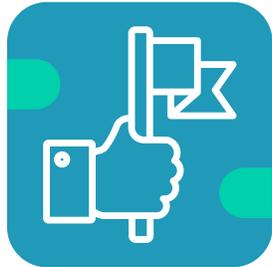
Es en este contexto del uso de los lenguajes computacionales, de la educación virtual y de las tecnologías digitales, un docente ubicado en la región del Bajo Cauca desarrolla la estrategia didáctica para ofrecer los cursos *Machine Learning* e *Introducción al Machine Learning* a los estudiantes de octavo, noveno y décimo semestre de los programas de Biología, Química, Estadística y Matemáticas de la Facultad de Ciencias Exactas y Naturales de la Universidad de Antioquia durante el semestre 2021-2.

Estos cursos se impartieron con la metodología de dinamización a través de la virtualidad, utilizando las plataformas Meet y Teams, y los cuadernos Jupyter como espacios de presentación de contenido y de experimentación activa de los mismos. Dicen Wu, Li, Chandru y Daniel que “los apoyos didácticos basados en multimedia son cada vez más populares en muchas disciplinas académicas” (2022, p. 2), y sabemos que en esos multimedios están incluidos todos los recursos digitales de las plataformas en Internet.

Si bien ante estas estrategias los estudiantes, por ser de semestre avanzados, tenían grandes conocimientos previos en el área de las asignaturas trabajadas en el aula, fue mediante las evaluaciones aplicadas en los cursos que se pudo percibir que las técnicas aportadas por el docente les generó inquietud y gran avidez, por lo que fácilmente pudieron articular esos saberes para desarrollar aún más sus habilidades como científicos de datos con capacidad para desenvolverse exitosamente en los retos laborales que exige el mundo industrial, científico y social.

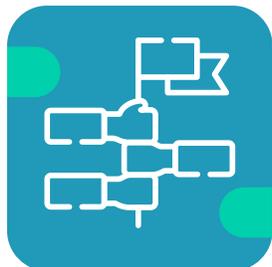
Igualmente, se espera que estos cursos puedan ser replicados en los programas regionalizados con el fin de brindar las mismas opciones de educación vanguardista a todos los estudiantes de la Universidad, independiente de la sede en donde se encuentren.

Objetivos de aprendizaje



Objetivo general

Desarrollar habilidades en ciencia de datos en el contexto de los estudiantes de los programas de Biología, Estadística, Química y Matemáticas de la Facultad de Ciencias Exactas y Naturales de la Universidad de Antioquia para la enseñanza y el aprendizaje de la ciencia de datos con el lenguaje Python.



Objetivos específicos

- Socializar el potencial del uso de herramientas tecnológicas, como los tutoriales de YouTube y los cuadernos Jupyter y de GitHub, en el aprendizaje de los estudiantes universitarios de semestres avanzados.
- Identificar cómo el uso del lenguaje Python facilita el análisis y la automatización de la información en los estudiantes de la Facultad de Ciencias Exactas y Naturales.
- Mejorar el conocimiento de las nuevas realidades, pero configurando una estructura de contenido interconectada que permita la interactividad del estudiante con los recursos de aprendizaje.



Resultados de aprendizaje

Como resultado de esta estrategia didáctica, los estudiantes:

- Desarrollan habilidades para obtener un dataset asociado a un problema investigativo o empresarial.
- Constituyen un modelo predictivo de Machine Learning para realizar una tarea de clasificación o de regresión, estimando su desempeño o poder de generalización.
- Están en capacidad de justificar la forma en que configuran su modelo.
- Presentan el proceso de su construcción y lo exponen con las interpretaciones de las predicciones, limitaciones y potencialidades que hacen del modelo propuesto.
- Podrán acceder a estos conocimientos desde otras sedes porque se espera que los productos de aprendizaje se puedan ofertar como cursos en los programas regionalizados.



La estructura conceptual estuvo enmarcada en la cuarta revolución industrial desde el big data, el Internet de las cosas (IoT), la automatización y la ciencia de datos que, en general, permiten el trabajo y el uso de diferentes técnicas y tecnologías en el aula de clase para la articulación del mundo real, social y empresarial con los aprendizajes de los profesionales en formación.

Se implementó la siguiente secuencia didáctica:



La metodología de presentación de contenidos se llevó a cabo a través de la constitución de una estructura de cuadernos Jupyter articulada con las redes sociales GitHub, YouTube, y Google Colab.



Viñals Blanco y Cuenca Amigo (2016) proponen que los nuevos roles del docente para la era digital son: organizador, guía, generador, acompañante, *coach*, gestor del aprendizaje, orientador, facilitador, tutor, dinamizador o asesor. De igual forma, establecen que estos roles cambian la idea de educación unidireccional por una horizontal, abundante, caótica y desestructurada, puesto que los modelos de enseñanza y aprendizaje no están centrados en el profesor sino en el estudiante.

Los docentes se enfrentan al reto de adquirir conocimientos para desarrollar las competencias que necesitan, saberes, habilidades y actitudes precisas para alcanzar los objetivos que se exigen desde el propio currículo formal (competencias digitales y aprender a aprender, entre otras), con el fin de adaptarse a las exigencias del mercado laboral y, aún más importante, descubrir sus verdaderas motivaciones, intereses e inquietudes. Estos roles van en consonancia con la concepción de aprendizaje como actividad experimental y activa.



Si bien existe una tendencia generalizada en la formación en ciencia de datos desde diversas universidades del mundo (Wu, Li, Chandru y Daniel, 2022; Fan y Zhong, 2022), y en construir repositorios de videos de apoyo a las clases diseñadas en cuadernos de Jupyter, nuestro factor diferenciador es que los videos del curso se articulan con los cuadernos y estos últimos son presentados al estudiante de forma dinámica para que pueda vincularlos con facilidad a plataformas como Google Colab, YouTube y GitHub, y leerlos de manera experimental, interactiva y activa. Además, el acceso a las notas de clase y a los bancos de ejemplos y de problemas es público, ya que la red social GitHub lo es, lo cual hace que el saber estructurado en esta estrategia didáctica sea de libre acceso y de carácter gratuito para cualquier estudiante. De esta forma, buscamos que la enseñanza de la ciencia de datos se vincule al programa curricular de cada carrera en la Universidad de Antioquia, dada su transversalidad en todas las áreas del conocimiento desde la docencia, la investigación y la extensión.

5 Evidencias



Dos repositorios en GitHub con las clases diseñadas, un banco de ejemplos y un banco de problemas:

- <https://github.com/marco-canas/intro-Machine-Learning>
- https://github.com/marco-canas/machine_learning



Un canal en YouTube llamado Dimathdata (Didáctica de la matemática y la ciencia de datos):

- <https://www.youtube.com/@MARDATA>



Videos de las exposiciones finales de los estudiantes sobre modelos en Machine Learning de regresión y clasificación.



La evaluación al docente, realizada por parte de los estudiantes, que refleja su aceptación de la estrategia didáctica aplicada en el aula.

6

Plan de evaluación



Lezcano y Vilanova (2017) recomiendan que la evaluación de formación en y para entornos virtuales debe tener una retroalimentación casi inmediata. Esto se tuvo en cuenta en una reunión que se llevó a cabo con los estudiantes, en el envío de las notas mediante correo electrónico, en las observaciones inmediatas realizadas a sus trabajos y en el reconocimiento de lo que, como docentes, aprendimos en la evaluación. Fue así como el diseño de las rúbricas para la evaluación de los trabajos finales de construcción de modelos de Machine Learning de clasificación y regresión estuvieron inspiradas por dichos autores.

Por tanto, en el plan de evaluación:

- Se realizaron tres quices y un parcial en línea con preguntas tipo ensayo y encuesta para evaluar la adquisición de los conceptos referentes al manejo del lenguaje Python para la ciencia de datos, específicamente, de las librerías NumPy, Pandas, Matplotlib y SciKit Learn.

La evaluación también incluyó dos componentes:

- Un primer componente teórico de medición de comprensión teórica del lenguaje Python y de los modelos lineales y no lineales del Machine Learning, y de comprensión

de las metodologías de preprocesamiento de datos, selección de modelos y afinamiento de estos.

- Un segundo componente de habilidad de codificación de los algoritmos de visualización de datos y de construcción de modelos de regresión y clasificación.

El trabajo de modelación final se llevó a cabo por medio de la exposición por equipos, donde se midieron las habilidades para comunicar el conocimiento, específicamente, la habilidad de comunicar los conocimientos referentes a la resolución de los problemas más frecuentes en aprendizaje supervisado.

Datos y evidencias

En el mundo globalizado actual es esencial que los estudiantes de diversas disciplinas, máxime los de matemáticas, estadística y las ciencias en general, desarrollen habilidades en la ciencia de datos, ya que habitan una sociedad cada vez más automatizada (Monleon-Getino y Canela-Soler, 2021). Por consiguiente, surge una didáctica de la ciencia de datos aplicada en la educación superior donde la sociedad del conocimiento, al ritmo de su evolución, impone tendencias en las que es necesario que el docente empoderado incursione en el diseño de contenidos teóricos y pedagógicos que combinen la tecnología, los lenguajes computacionales, las plataformas digitales de la Inteligencia Artificial, el Internet de las cosas, el big data, la minería de datos y todos los demás requerimientos necesarios para ofrecer una educación audaz y acertada en el desarrollo de habilidades cognitivas, competencias comunicacionales, y análisis y transformación de la información con los estándares que exige el nuevo modelo de trabajo de la cuarta revolución (Mendoza Jurado, 2018).

De acuerdo con lo anterior, en la estrategia desarrollada con los estudiantes de ambos cursos se puede evidenciar lo siguiente:

- Los estudiantes manifestaron el uso de los videos de apoyo del canal de YouTube del curso.
- Los estudiantes manifestaron el uso del vínculo Open in Colab que les permitía articular el repositorio de las clases en GitHub con Google Colab, lo cual les facilitó una lectura experimental y activa de las clases.

- El material audiovisual y digital de apoyo que los estudiantes recopilaron en las distintas plataformas obtuvo un gran número de visitas y reproducciones (60 visitas en solo tres días después de publicado), estrategia que hace que se continúe, en el tiempo, con los aprendizajes de los contenidos ofrecidos en el aula virtual.

A continuación, mostramos algunos de los resultados de aprendizaje de los estudiantes:



Resultados de aprendizaje de los estudiantes

- Cuaderno de Yeifri visualizable a través de GitHub [aquí](#).
- Cuaderno Merlis sobre modelación [aquí](#).
- Exposición de Jose Luis Camilo sobre su modelo de reconocimiento de imágenes de tres tipos de moléculas [aquí](#).



Factores críticos



El factor crítico más importante es la oferta curricular de cursos de Machine Learning u otros afines a la ciencia de datos, transversales a diferentes programas y que, por su quehacer profesional, se requieren para el buen desempeño empresarial, científico y académico en el mundo actual. Este factor también es replicable en los programas regionalizados donde el curso habría que ajustarse curricularmente, de acuerdo con las tendencias globales de la ciencia y la tecnología que demanda la enseñanza en la educación superior. Así lo explicita el Plan de Desarrollo Institucional 2017-2027 y el Plan de Acción 2021-2024 de la Universidad de Antioquia cuando se menciona la evolución de los procesos de formación con equidad e inclusión.

En consecuencia, la enseñanza de la ciencia de datos, como tendencia universal en la educación superior, es transversal a todas las áreas del conocimiento, por lo que se pretende que esta estrategia innovadora se incluya como asignatura en los programas de las licenciaturas en educación, ingenierías, y ciencias exactas y naturales que tienen presencia en la región, con el fin de que los estudiantes de las distintas sedes salgan al mundo profesional, empresarial e industrial con las mismas capacidades y oportunidades de aquellos que estudian en Medellín.

Referencias bibliográficas

- Fan, X. y Zhong, X. (2022). Artificial intelligence-based creative thinking skill analysis model using human-computer interaction in art design teaching. *Computers and Electrical Engineering*, 110, 107957. <https://doi.org/10.1016/j.compeleceng.2022.107957>
- Lezcano, L. y Vilanova, G. (2017). Instrumentos de evaluación de aprendizaje en entornos virtuales. Perspectiva de estudiantes y aportes de docentes. *Informes Científicos Técnicos-UNPA*, 9(1), 1-36. <https://publicaciones.unpa.edu.ar/index.php/ICTUNPA/article/view/560>
- Mendoza Jurado, H. F. (2018). Ciencia de datos una alternativa de análisis al crecimiento pedagógico del estudiante en educación superior. *Educación Superior*, 5(2), 35-46. http://www.scielo.org.bo/scielo.php?pid=S2518-82832018000200006&script=sci_abstract
- Mitchell, T. et al. (1997). *Machine Learning: An Artificial Intelligence Approach* (Volume I). Morgan Kaufmann.
- Monleon-Getino, A. y Canela-Soler, J., (2021). La llegada de la ciencia de los datos en el entorno de la docencia universitaria de las ciencias de la vida y de la salud: propuestas para la adaptación de las asignaturas de matemáticas, estadística y análisis de datos. En Romero García, C. y Buzón García, O. (2021). *Innovación e investigación docente en educación: experiencias prácticas*. Dykinson. <https://bit.ly/3FATf0S>
- Salas-Rueda, R. A., Ramírez-Ortega, J., Eslava-Cervantes, A. L., Castañeda-Martínez, R. y De-La-Cruz-Martínez, G. (2022). Percepción de los profesores sobre los juegos web y dispositivos móviles en el nivel educativo superior durante la pandemia COVID-19. *Texto libre*, 15. <https://doi-org.udea.lookproxy.com/10.35699/1983-3652.2022.37074>

- Schwab, K. M. (2016). *La cuarta Revolución industrial*. World Economic Forum. [http://40.70.207.114/documentosV2/La%20cuarta%20revolucion%20industrial-Klaus%20Schwab%20\(1\).pdf](http://40.70.207.114/documentosV2/La%20cuarta%20revolucion%20industrial-Klaus%20Schwab%20(1).pdf)
- Universidad de Antioquia. (2017). *Plan de desarrollo 2017-2027: Una universidad innovadora para la transformación de los territorios*. <http://www2.udea.edu.co/webmaster/multimedia/plan-desarrollo-udea/plan-desarrollo-udea.pdf>
- Universidad de Antioquia. (2021). *Plan de Acción Institucional 2021-2024: Una universidad solidaria comprometida con la vida, la equidad y la diversidad, con visión global y pertinente frente a los retos de la sociedad*. <https://bit.ly/3wpr6pk>
- Viñals Blanco, A. y Cuenca Amigo, J. (2016). *El rol del docente en la era digital*. <https://bit.ly/3L7HGzo>
- Wu, K., Li, C. y Daniel, A. (2022). Digital teaching in the context of Chinese universities and their impact on students for Ubiquitous Applications. *Computers and Electrical Engineering*, 100, 107951. <https://doi.org/10.1016/j.compeleceng.2022.107951>