

DATOS EDUCATIVOS: PRODUCCIÓN Y USO DE HERRAMIENTAS DE PROCESAMIENTO DE INFORMACIÓN CON ASISTENCIA DE LA IA PARA LA TOMA DE DECISIONES

FERNANDO JORGE BIFANO

Facultad de Ciencias Exactas y Naturales — Universidad de Buenos Aires

`fjbifano@ccpems.exactas.uba.ar`

PABLO FABIÁN CARRANZA

Universidad Nacional de Río Negro

RESUMEN. La generación y uso de datos es actualmente uno de los nuevos emergentes que impactan en las diferentes esferas de la sociedad y por tanto, la educación se ve tensionada por las demandas que ello conlleva. En ese sentido, surgen nuevas vacancias para la formación docente y se vuelve necesario ofrecer instancias que permitan a quienes enseñan contar con herramientas para la comprensión, uso y toma de decisiones fundadas y sostenidas en la interpretación de la información. Este taller ofrece la posibilidad de familiarizarse con herramientas básicas que brindan los softwares para el procesamiento de la información y que con la asistencia de las IA se ven potenciados en sus aplicaciones. A partir del trabajo con distintos tipos de bases de datos educativos, propondremos un recorrido que aborda algunas herramientas de representación gráfica de la información en lenguaje de programación Python asistido por medio de inteligencias artificiales tales como ChatGPT y Gemini. Se analizarán las potencialidades de estrategias tales como clustering y árboles de decisión para la caracterización de grupos de estudiantes.

PALABRAS CLAVE — Datos educativos, IA, Python, Procesamiento de la información.

1. INTRODUCCIÓN

El análisis de datos es una disciplina que ha visto incrementado su desarrollo recientemente por varios factores vinculados entre sí. Entre ellos destacamos:

- Disponibilidad de herramientas de procesamiento de datos en computadoras de escritorio.
- Crecimiento de librerías de libre acceso en lenguaje Python.
- Desarrollo de nuevos métodos de análisis de datos, llegando incluso a herramientas vinculadas a la inteligencia artificial.
- Existencia de volúmenes de datos.

Estos elementos combinados permiten la aplicación de herramientas que facilitan la comprensión de fenómenos complejos y la toma de decisiones basadas en información.

Las instituciones educativas no escapan a estos avances. En efecto, los datos disponibles o factibles de ser obtenidos permiten una mejor interpretación de fenómenos del ecosistema educativo y el desarrollo de acciones basadas en información. A modo de ejemplo citamos:

- análisis de desempeño integral de estudiantes
- relaciones entre estrategias de estudio y calificaciones
- relaciones entre proyectos de vida, resiliencia y retención universitaria
- sistemas de alerta temprana de abandono
- caracterización de grupos de estudiantes por técnicas de clustering

El análisis de datos, entonces, resulta una herramienta que permite extraer información para la comprensión de fenómenos que acontecen en la institución educativa y así tomar decisiones basadas en información precisa y contextualizada.

Estas herramientas pueden ser utilizadas tanto por los equipos de conducción como por los docentes en la aulas; en contextos presenciales o virtuales. En esta propuesta nos centraremos en algunas herramientas de análisis de datos que resultan de interés a docentes tanto sea al interior de una cátedra como para el análisis interdisciplinario del desempeño de estudiantes. Las mismas son también de gran interés para la gestión de la institución, facilitando la comprensión de la dinámica de la misma y brindando fundamentos para la toma de decisiones.

2. CONTENIDOS

Se abordarán algunas herramientas de representación gráfica de la información en lenguaje de programación Python asistido por medio de inteligencias artificiales tales como ChatGPT y Gemini. Se analizarán las potencialidades de herramientas tales como clustering y árboles de decisión para la caracterización de grupos de estudiantes. Más precisamente se proponen los siguientes módulos:

2.1 Actividades previas

- a) Introducción a lenguaje de programación Python en Jupyter Notebook
El análisis de datos se realizará en el entorno llamado Jupyter Notebook, herramienta accesible desde el software Anaconda (gratuito).
- b) Abrir una base de datos en formatos xlsx o csv en Jupyter Notebook con Python
Los datos serán analizados en python, por lo que el primer paso será subir la base de datos al entorno Jupyter Notebook para poder tratar la base de datos en el lenguaje python.

2.2 Módulo 1

En este módulo se abordarán los conceptos básicos de programación en Python en Jupyter Notebook para un primer tratamiento de datos educativos. En esta primera etapa, se trabajará con datos numéricos reales anonimizados de alumnos de escuelas secundarias de Argentina. Entre otros, serán tratados temas tales como:

- a) primera exploración de la base de datos
- b) representación gráfica de variables
- c) correlaciones
- d) normalización o estandarización
- e) clustering por método k-nn
- f) árboles de decisión

Todos los métodos serán considerados como herramientas de síntesis para la extracción de información de los datos. En todas las instancias se accederá al uso de IA generativas como asistentes a la programación en Python.

En este módulo se debatirá sobre las condiciones del trabajo final del taller, consistente en una presentación por dupla de un análisis de datos propios. Más precisamente, se consagrarán los últimos minutos a establecer las características que deben reunir los datos para un análisis acorde a las posibilidades del taller.

2.3 Módulo 2

En este módulo se abordará el tratamiento de datos categóricos en escala de likert, muy frecuentes en contextos educativos. Se tratarán en principio con los mismos métodos abordados en el módulo 1. Se integrarán también con datos de origen numérico para posibles extracciones de información de tales cruzamientos.

Se dedicarán los últimos minutos a terminar de definir las características de la presentación que cada dupla hará sobre el análisis de datos propios como trabajo final.

2.4 Módulo 3

Este módulo se consagrará enteramente a las presentaciones que cada dupla haya podido realizar sobre el análisis de datos educativos propios. Se realizarán también devoluciones y comentarios de parte de todos los participantes a las diferentes presentaciones.

3. REQUISITOS PREVIOS

Este taller está destinado a docentes que enseñan matemáticas en los niveles medio y superior. Se requiere conocimientos mínimos en el área de la estadística descriptiva. Así mismo resulta deseable que los asistentes cuenten con cierta familiaridad en relación con el uso de herramientas tecnológicas tales como Excel. El manejo de lenguaje de programación previo no es excluyente y se espera que sea una de las capacidades a desarrollar en los asistentes.

4. OBJETIVOS

En este taller, nos proponemos alcanzar los siguientes objetivos:

- Favorecer la reflexión crítica sobre la producción y el uso de datos educativos para la toma de decisiones fundamentadas.
- Introducir en el uso de los rudimentos básicos de la programación en Python para el análisis de datos.
- Utilizar herramientas tecnológicas, especialmente con la asistencia de la IA, para potenciar el análisis de la información educativa para la toma de decisiones.
- Desarrollar habilidades básicas en el tratamiento de bases de datos para el análisis estadístico.

5. ACTIVIDADES

5.1 Actividades previas

Para poder participar plenamente del taller, les proponemos a los cursantes que como actividades previas se familiaricen con algunos softwares y/o herramientas que utilizaremos durante el desarrollo del mismo. Específicamente les proponemos:

- Para una introducción al tema sugerimos seguir las propuestas detalladas en los siguientes vídeos:
Descargar Anaconda: <https://www.anaconda.com/download>
Vídeos de primeros pasos de Python en Jupyter Notebook:
 - <https://www.youtube.com/watch?v=81-AyuxjdBo>
 - <https://www.youtube.com/watch?v=cp5gUF1D0nQ&list=PLMUoURdFUxkkHw8tVweJi8YAZcZHDjYou>
- Para una introducción a la etapa de acceso a una base de datos desde Jupyter notebook, los participantes pueden acceder al siguiente vídeo:
 - <https://www.youtube.com/watch?v=dbEQtzObsQw>Los enlaces no son excluyentes, los participantes encontrarán muchos otros en la web.

5.2 Primera hora y media sincrónica

A lo largo de la primera etapa sincrónica, le propondremos a los cursantes, el análisis de una base de datos anonimizada basada en resultados de rendimiento académico (módulo 1) —provista por los profesores a cargo del taller— que se centrará en aspectos tales como: representación gráfica de datos, correlaciones, clustering, entre otros.

- Exploración de base de datos.
- Representación gráfica de distribuciones de variables.
- Establecimiento de posibles relaciones entre variables.
- Utilización de métodos de machine learning tales como clustering y árboles de decisión.

5.3 Primeras tres horas entre clases

Como parte de la primera sección entre clases, los estudiantes en grupos deberán producir un breve informe con los principales hallazgos obtenidos a partir del análisis estadístico efectuado en la sección 5.2. Las características del mismo son:

- Una extensión máxima de 5 carillas incluyendo gráfico y/o anexos.
- Expresar clara y fundadamente, los principales resultados que surgen del análisis estadístico de la base de datos estudiada.

5.4 Segundas hora y media sincrónicas

La primera parte de la sesión estará abierta a compartir algunos de los principales resultados producidos en los informes grupales como fruto de la etapa 5.3. En una segunda parte, análogamente a lo descrito en la sección 5.2, los profesores del taller ofreceremos a los cursantes una base de datos educativos relacionados con el abandono escolar para su análisis (módulo 2).

5.5 Segundas tres horas entre clases

Este espacio es análogo al descrito en la sección 5.3. La diferencia fundamental radica en que se les propondrá a los cursantes del taller, la búsqueda y/o elaboración de información que pueda ser un insumo para la construcción de una base de datos propia que les permita hacer un estudio de alguna problemática relacionada con las instituciones escolares en la que se desempeñan. Esto será el insumo fundamental para la producción final.

5.6 Terceras hora y media sincrónicas

Este espacio es análogo al descrito en la sección 5.4. A diferencia de la etapa referenciada, dedicaremos una primera parte de la sesión a que los estudiantes compartan los datos recogidos en la etapa 5.5. a los fines de discutir la viabilidad de su estudio para la producción final (módulo 3).

5.7 Evaluación final

Como parte de la evaluación final del curso, propondremos a los cursantes las siguientes actividades:

- Elaboración de un proyecto final aplicando los conocimientos adquiridos. Más precisiones sobre el mismo se ofrecerán oportunamente a través de la plataforma de la facultad.
- Cuestionario final que aborde los principales temas del taller a modo de autoevaluación del taller.
- Reflexión escrita sobre la experiencia del taller y el aprendizaje obtenido dando cuenta de los principales obstáculos superados a lo largo de la experiencia.

6. BIBLIOGRAFÍA

- [9] Aristizabal, J. (2016). Analítica de datos de aprendizaje (ADA) y gestión educativa. *Rev. Gestión de la Educación*, 6(2), 149-168. <https://doi.org/10.15517/rge.v1i2.25499>
- [10] Barragán Pazmiño, B. M., & Pazmiño Maji, R. A. (2018). Literatura científica sobre análisis estadístico implicative: un mapeo sistemático de la década que transcurre. *Ciencia Digital*, 2(4.1), 55-69. <https://doi.org/10.33262/cienciadigital.v2i4.1.190>
- [11] Carranza, P., Vasches, F., Gamoneda, A., Cardieri, M., Benzaquén, I., & Pedersen, E. (2024). Estudio sobre violencia en relaciones sexoafectivas de jóvenes de la provincia de Neuquén, Argentina mediante A.S.I. <http://rid.unrn.edu.ar/handle/20.500.12049/8919>
- [12] Carranza, P. (2014). Presencia de interpretaciones bayesiana y frecuentista de la probabilidad en libros de estudio en Francia [Especial ASI]. *Educ. Matem. Pesq., São Paulo*, 16(3), 1071-1087.
- [13] Fajardo Ibarra, R., & González Zúñiga, F. (2017). El impacto de los videojuegos en la inteligencia: Una revisión de la literatura. *Innovación Educativa*, 17(74), 131-149. <https://doi.org/10.22458/ie.v17i74.25499>
- [14] Gámez Jara, D., & Rodríguez Martínez, L. (2016). La influencia de la tecnología en el rendimiento académico de los estudiantes de secundaria. *Revista Latinoamericana de Tecnología Educativa*, 15(2), 30-50. <https://doi.org/10.22458/rte.v15i2.25499>
- [15] Morduchowicz, A. (2006). Los indicadores educativos y las dimensiones que los integran. <http://biblioteca.udgvirtual.udg.mx/jspui/handle/123456789/198>
- [16] UNESCO IIEP Learning Portal. (2024). Usar datos para mejorar la calidad. <https://learningportal.iiep.unesco.org/es/fichas-praticas/monitorear-el-aprendizaje/usar-datos-para-mejorar-la-calidad>