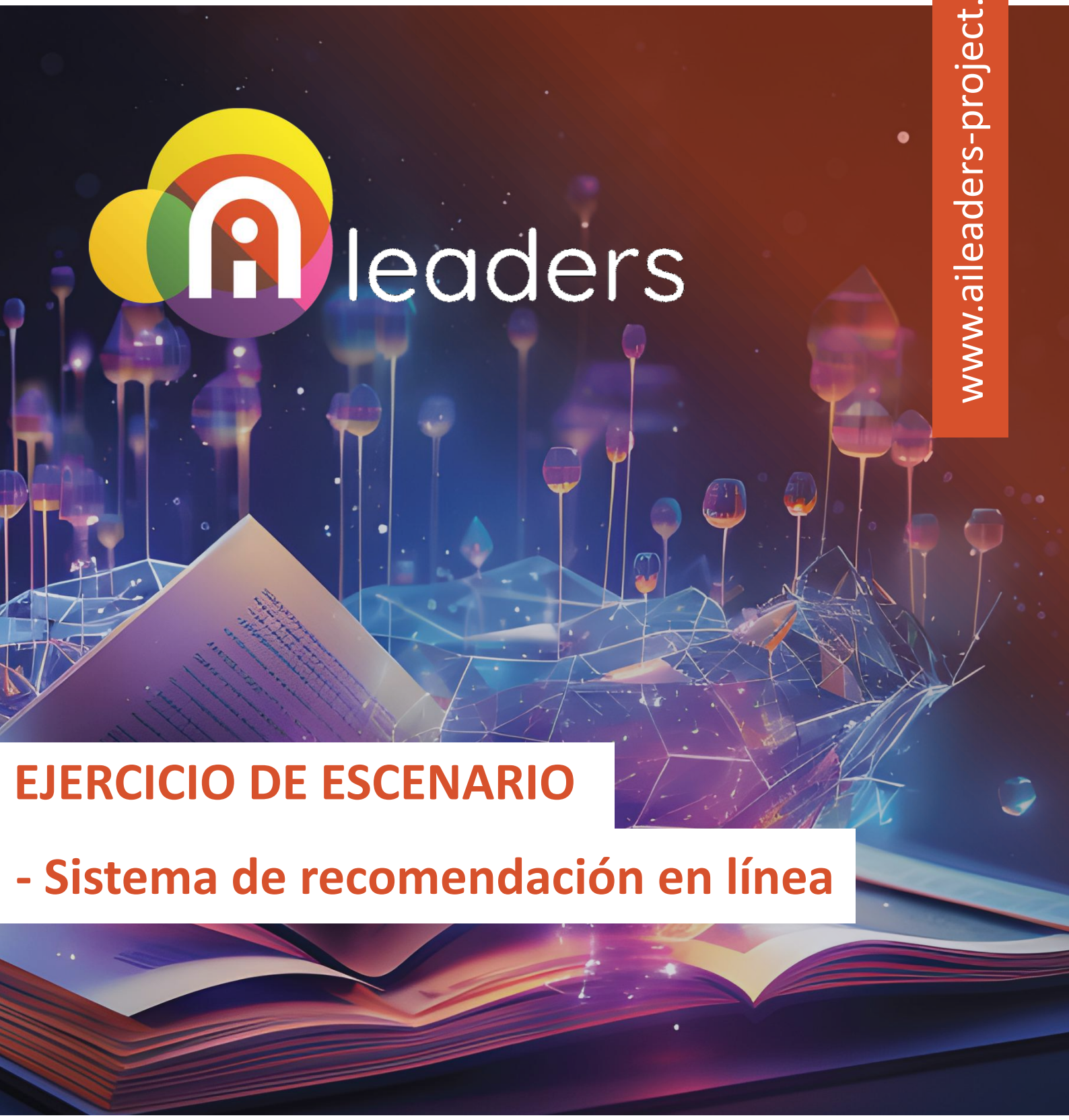


www.aileaders-project.eu

EJERCICIO DE ESCENARIO

- Sistema de recomendación en línea

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

EJERCICIO DE ESCENARIO
Sistema de recomendación en línea

01 Resumen

Tipo de REA

Ejercicio de escenario utilizando Google Colab (riverML)

Objetivo/Finalidad

Este ejercicio basado en un escenario sitúa a los alumnos en el papel de un profesional del aprendizaje automático (ML) encargado de analizar la imparcialidad de un sistema de recomendación en línea. Explora cómo la adaptación continua al comportamiento de los usuarios puede amplificar los sesgos existentes con el tiempo. Los alumnos tendrán el reto de identificar los sesgos y evaluar estrategias para mitigarlos.

Resultados de aprendizaje esperados

El estudiante será capaz de:

- Identificar posibles fuentes de sesgo en las recomendaciones generadas por ML.
- Comparar métricas de equidad, como la paridad demográfica y la equidad de exposición.
- Comprender cómo los bucles de retroalimentación afectan a la equidad a lo largo del tiempo en los sistemas de recomendación en línea.

Palabras clave

- Aprendizaje automático
- Recomendación
- Sesgos
- Equidad

Enfoque: Metodológico Sugerido

Aprendizaje basado en problemas.

02 Introducción

En el contexto de las experiencias digitales cada vez más personalizadas, los sistemas de recomendación basados en la inteligencia artificial desempeñan un papel fundamental a la hora de moldear el comportamiento de los consumidores, influir en las decisiones de compra y determinar la visibilidad de los productos y contenidos en los distintos segmentos de usuarios.

Si bien estos sistemas ofrecen ventajas considerables para la optimización del marketing y los negocios, también plantean preocupaciones éticas y normativas apremiantes, en particular en lo que respecta a la equidad, la transparencia y la posible discriminación (Akter et al., 2022; Bozdag, 2013).

Los sistemas de recomendación no son meros artefactos algorítmicos, sino construcciones sociotécnicas complejas, influenciadas por decisiones humanas en cada etapa, desde la recopilación de datos y el diseño de modelos hasta la interpretación y el despliegue de los resultados (Bozdag, 2013). Los sesgos pueden surgir de múltiples fuentes: prejuicios humanos, limitaciones técnicas o desajustes contextuales. Especialmente en entornos de marketing, estos sesgos pueden interactuar y agravarse, configurando no solo el rendimiento del sistema, sino también las experiencias de los usuarios y los resultados sociales (Akter et al., 2022).

Una preocupación importante radica en la naturaleza iterativa de los sistemas de recomendación, que dependen de los comentarios de los usuarios para reentrenar y perfeccionar los modelos predictivos. Este bucle de retroalimentación puede reforzar los desequilibrios existentes entre la entrada y la salida, lo que conduce a disparidades persistentes a lo largo del tiempo (Sun et al., 2020). A medida que los modelos incorporan continuamente nuevos datos y preferencias de los usuarios, su equidad y precisión están sujetas a cambios dinámicos. Si no se controla, esto puede dar lugar a un sobreajuste a los comportamientos dominantes de los usuarios, lo que exacerba la infrarrepresentación de los grupos minoritarios y restringe la diversidad de contenidos, un fenómeno asociado al control algorítmico y al auge de las burbujas de filtro (Harambam et al., 2018).

Motivado por estos retos, este ejercicio basado en escenarios invita a los estudiantes a explorar las dimensiones éticas de los sistemas de recomendación utilizando el **conjunto de datos MovieLens 100K** como banco de pruebas.

A través de la experimentación práctica, los estudiantes examinan cómo los bucles de retroalimentación algorítmica pueden amplificar involuntariamente las desigualdades, en particular al sesgar la visibilidad del contenido y favorecer las preferencias de la mayoría. Se les anima a investigar tanto los aspectos técnicos como los éticos de estos sistemas, tendiendo un puente entre los conocimientos teóricos y el análisis práctico. Este enfoque pedagógico fomenta la conciencia crítica sobre los riesgos asociados a la personalización poco ética, entre los que se incluyen la pérdida de confianza de los consumidores, el daño a la reputación y la disminución del acceso a oportunidades para los usuarios marginados. El ejercicio pone de relieve que la equidad en la IA no es una evaluación estática, sino una responsabilidad dinámica y continua. A medida que evolucionan las preferencias de los usuarios y cambia el contenido, los modelos también deben adaptarse para mantener un rendimiento equilibrado en todos los grupos de usuarios. En última instancia, esta experiencia de aprendizaje dota a los estudiantes de las herramientas necesarias para diseñar sistemas de IA que no solo sean eficaces y eficientes, sino también equitativos, transparentes y socialmente responsables.

Métricas de equidad

01 **Error de estimación absoluto**: mide el error absoluto entre la calificación prevista y la calificación real dada por el usuario u al elemento I, definido por:

$$\varepsilon_a(u,i) = |\hat{r}_{u,i} - r_{u,i}|$$

02 **Error de sobreestimación**: mide en qué medida la predicción sobreestima la calificación real, definido por:

$$\varepsilon_o(u,i) = \max(\hat{r}_{u,i} - r_{u,i}, 0)$$

03 **Error de subestimación**: mide en qué medida la predicción subestima la calificación real, definido por:

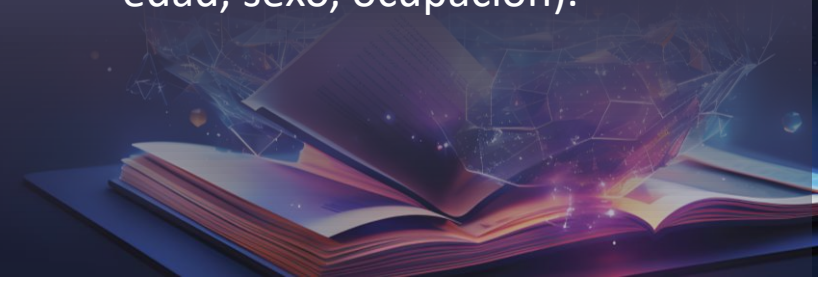
$$\varepsilon_u(u,i) = \max(r_{u,i} - \hat{r}_{u,i}, 0)$$

04 **Error de estimación del valor**: mide el error con signo entre la calificación prevista y la calificación real, definido por:

$$\varepsilon_v(u,i) = \hat{r}_{u,i} - r_{u,i}$$

03 Presentación de herramientas

Este ejercicio aprovecha el **conjunto de datos MovieLens 100K**, un punto de referencia ampliamente utilizado en la investigación de sistemas de recomendación. El conjunto de datos incluye 100 000 valoraciones, 943 usuarios, 1682 películas, junto con datos demográficos de los usuarios (por ejemplo, edad, sexo, ocupación).



Aunque es útil para probar algoritmos el conjunto de datos contiene importantes desafíos de imparcialidad

01 **Representación desequilibrada entre los grupos demográficos;**

02 **Distribuciones sesgadas de las calificaciones;**

03 **Representación excesiva de los artículos populares.**

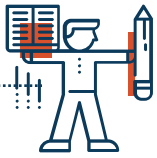
Estas características hacen que MovieLens 100K sea ideal para explorar el sesgo y la imparcialidad en los algoritmos de recomendación.

RiverML[1] es una biblioteca Python de código abierto diseñada específicamente para el aprendizaje automático en flujos de datos. Proporciona herramientas para crear, entrenar y evaluar modelos que aprenden de forma incremental a partir de datos que llegan en tiempo real, en lugar de por lotes. RiverML es compatible con algoritmos de aprendizaje supervisado y no supervisado, y ofrece funciones como el aprendizaje adaptativo, la detección de desviaciones conceptuales y el preprocesamiento de flujos. Su diseño modular facilita la combinación de modelos, pasos de preprocesamiento y estrategias de evaluación, lo que lo convierte en una opción potente para aplicaciones de aprendizaje en línea, como la detección de fraudes, los sistemas de recomendación y el análisis de datos de sensores. Las métricas de equidad disponibles en la biblioteca fueron implementadas por el equipo de FEP, lo que contribuyó a la evaluación ética de los modelos de streaming.

Esta herramienta permite la ejecución de procesos de evaluación de la equidad y la comparación visual del rendimiento de las recomendaciones entre grupos y a lo largo del tiempo.

1. Tutorial de Riverml: https://riverml.xyz/latest/introduction/installation/

04 Actividades prácticas



Actividad

01 Acceda al cuaderno Jupyter Notebook proporcionado en https://tinyurl.com/5n8vvpx9

02 Ejecute todas las celdas del cuaderno (haga clic en la pestaña «Runtime» y, a continuación, en «Run All»).

03 Analice y compare las métricas de equidad entre diferentes grupos de usuarios (por ejemplo, por género o edad).

04 Observe cómo evolucionan los indicadores de equidad a medida que se recopilan más interacciones de los usuarios.

Preguntas para el debate

01 ¿Qué grupos de usuarios se ven sistemáticamente favorecidos o perjudicados?

02 ¿Cómo cambia la calidad de las recomendaciones con el tiempo?

03 ¿Qué compensaciones surgen al intentar equilibrar el compromiso con la equidad?

05 Conclusión

Este ejercicio ilustra la naturaleza dinámica de la equidad en los sistemas de recomendación en línea.

Aunque el rendimiento inicial del sistema puede parecer imparcial, su funcionamiento a largo plazo puede dar lugar a disparidades significativas debido a los bucles de retroalimentación y al sesgo de popularidad. Estos desequilibrios afectan a la experiencia y la confianza de los usuarios, especialmente en el caso de los grupos de usuarios infrarrepresentados o minoritarios. Al aplicar métricas de equidad y analizar la evolución del sesgo en tiempo real, los estudiantes adquieren conocimientos prácticos sobre el despliegue ético del aprendizaje automático. También exploran las limitaciones de la evaluación estática y la importancia de la supervisión continua de la equidad en los sistemas adaptativos. El ejercicio subraya la tensión entre la optimización del compromiso y la garantía de un trato equitativo, un reto clave para el diseño ético de la IA.

06 Referencias

- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. Ethics and Information Technology, 15(3), 209–227.
- Akter, S., Michael, K., Bandara, R., Wamba, S.F., Foropon, C., & Papadopoulos, T. (2022). Algorithmic bias in machine learning–based marketing models: A dynamic capability view. Journal of Business Research.
- Harambam, J., Helberger, N., & van Hoboken, J. (2018). Democratizing algorithmic news recommenders: How to materialize voice in a technologically saturated media ecosystem. Big Data & Society.
- Sun, T., Gautam, K., & Joachims, T. (2020). Debiasing Learning from User Interaction. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.

Sigue nuestro viaje

www.aileaders-project.eu

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.