



www.aileaders-project.eu

SIMULACIÓN - Retos éticos

**en recursos
humanos**



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.

SIMULACIÓN

- Retos éticos en recursos humanos

01



Resumen 3

02



Introducción 4

03



Presentación de herramientas 6

04



Ejecución de la simulación 7

05



Conclusiones 8

06



Referencias 9

• 01 Resumen



Tipo de REA

Demo/simulación con Google Colab (clasificador RandomForest frente a FairGBM).

Objetivo/Finalidad

Proporcionar a los alumnos una exploración práctica y crítica de cómo la toma de decisiones algorítmica en la contratación puede reproducir desigualdades estructurales mediante la comparación de los resultados de las métricas de equidad. Esta simulación fomenta la reflexión sobre las dimensiones éticas de la IA en los recursos humanos y promueve el desarrollo de modelos predictivos más justos e inclusivos.

Resultados de aprendizaje esperados

Al final de la simulación, el alumnado será capaz de:

- 01** Detectar e interpretar el sesgo algorítmico en la contratación impulsada por la IA.
- 02** Reflexionar sobre las implicaciones éticas de los sistemas de contratación automatizados.
- 03** Utilizar métricas de equidad para evaluar los resultados del modelo.

Palabras clave

- Aprendizaje automático
- Clasificación
- Recursos humanos
- Sesgos
- Equidad

Enfoque metodológico sugerido:

Aprendizaje basado en problemas

• 02 Introducción



Glosario

- 01** **La responsabilidad** significa que la organización es responsable de sus decisiones de contratación y sus resultados, garantizando que existan consecuencias para las prácticas injustas.
- 02** **La transparencia** implica que el proceso de contratación y los criterios de selección sean claros y comprensibles para todos los candidatos.
- 03** **La equidad** garantiza que todos los candidatos sean evaluados de manera imparcial en función de sus cualificaciones para el puesto, sin sesgos ni discriminaciones.
- 04** **El sesgo social** en la contratación es la tendencia injusta a favorecer o rechazar a los candidatos basándose en suposiciones prejuiciosas sobre su grupo social (como la edad, el género o la raza) en lugar de sus habilidades y cualificaciones reales.

Cada vez más se espera que los procesos de selección cuenten con el apoyo de tecnologías de inteligencia artificial (IA), especialmente en las primeras etapas de la contratación, como la selección de currículos, la clasificación de candidatos y la preselección de solicitantes para las entrevistas. Si bien estos sistemas ofrecen un mayor potencial de eficiencia y escalabilidad, también plantean importantes cuestiones éticas, en particular en lo que respecta a la equidad y los prejuicios (Horodyski, 2023).



Mucho antes de la adopción de la IA, las investigaciones sobre las prácticas tradicionales de contratación y selección ya habían documentado patrones persistentes de discriminación (Breaugh, 2013; Hebl et al., 2020). Entre ellos se incluyen los sesgos relacionados con la raza, el origen étnico y la condición de minoría (por ejemplo, Allen y Vardaman, 2017; Hiemstra et al., 2013; Veit y Thijssen, 2019; Zschirnt y Ruedin, 2016); género (por ejemplo, Ellemers, 2018); clase social (por ejemplo, Henderson, 2018); edad (por ejemplo, Czopp et al., 2015; Zaniboni et al., 2019); y formación académica (por ejemplo, Daly et al.,

2000).

Cada vez más, se espera que los procesos de selección cuenten con el apoyo de tecnologías de inteligencia artificial (IA), especialmente en las primeras etapas de la contratación, como la selección de currículos, la clasificación de candidatos y la preselección de solicitantes para entrevistas, lo que puede plantear importantes cuestiones éticas, en particular en lo que respecta a la equidad y los prejuicios (Horodyski, 2023).

Mucho antes de la adopción de la IA, las investigaciones sobre las prácticas tradicionales de contratación y selección ya habían documentado patrones persistentes de discriminación (Hebl et al., 2020). Entre ellos se incluyen los prejuicios relacionados con la raza, la edad, el origen étnico y la condición de minoría (por ejemplo, Allen y Vardaman, 2017; Veit y Thijsen, 2019; Zaniboni et al., 2019).



La preocupación ahora es que las tecnologías de IA, en lugar de mitigar estos sesgos, puedan afianzarlos o incluso amplificarlos. Estudios recientes sugieren que la toma de decisiones algorítmica en recursos humanos puede replicar las desigualdades históricas incorporadas en los datos utilizados para entrenar dichos sistemas (por ejemplo, Rigotti y Fosh-Villaronga, 2024; Seppälä y Małecka, 2024). Como resultado, el uso de la IA en la contratación plantea cuestiones críticas sobre la responsabilidad, la transparencia y la equidad en las prácticas de contratación.

En este contexto, el presente ejercicio de simulación está diseñado para involucrar a los estudiantes en los retos éticos de la contratación y la selección de candidatos impulsadas por la IA. El objetivo es proporcionar experiencia práctica en la identificación de posibles sesgos dentro de los algoritmos de contratación y explorar estrategias

conscientes de la equidad para apoyar una toma de decisiones más inclusiva y equitativa. Mediante el uso de métricas de equidad, los estudiantes analizarán si los modelos de aprendizaje automático utilizados para predecir las decisiones de preselección reflejan —o reproducen— patrones históricos de discriminación, prestando especial atención al origen étnico, el género y el contexto socioeconómico.

Al enfrentarse a estos retos en un entorno controlado y simulado, los estudiantes desarrollarán tanto la competencia técnica como la sensibilidad ética esenciales para el diseño y la implementación responsables de la IA en la gestión de recursos humanos.



• 03 Presentación de herramientas



Esta simulación aborda una tarea de clasificación binaria en la contratación basada en IA: predecir si un candidato debe ser preseleccionado para un puesto directivo de nivel medio en función de su perfil.

El conjunto de datos utilizado es sintético, pero se basa en datos de contratación del mundo real, disponibles públicamente a través de la plataforma Kaggle. Los atributos de los candidatos incluyen género, raza/etnia, nivel de educación, años de experiencia, expectativas salariales y fuente de contratación.

Aunque el conjunto de datos es artificial, simula de forma realista las decisiones y los patrones comunes de contratación, lo que lo hace ideal para estudiar la equidad en el aprendizaje automático. Es especialmente preocupante la forma en que variables sensibles como la raza/etnia pueden influir en las predicciones del modelo. Los estudios

han demostrado que los procesos de contratación pueden perjudicar sistemáticamente a los candidatos de determinados orígenes étnicos, incluso cuando las cualificaciones son iguales (Zschirnt y Ruedin, 2016; Veit y Thijsen, 2019; Hiemstra et al., 2013).

Esta simulación se centra en una **pregunta principal**:
¿Los algoritmos de aprendizaje automático «recomiendan» de manera desproporcionada el rechazo de las solicitudes de empleo presentadas por personas negras y mujeres?

Al investigar esta cuestión, los estudiantes evalúan de forma crítica si los modelos de IA utilizados en la selección de candidatos muestran sesgos no deseados y cómo estos patrones se relacionan con los hallazgos empíricos de la literatura académica.

• 04 Ejecución de la simulación



01 Acceda al cuaderno de simulación

Vaya a <https://tinyurl.com/k63793wp>

02 Ejecutar el código

- Para ejecutar todas las celdas, haga clic en la pestaña «Runtime» y seleccione «Run All». Espere un momento (puede tardar unos pocos minutos).
- Asegúrese de que los resultados se cargan correctamente y de que todos los modelos se han entrenado con éxito.

03 Explore el conjunto de datos

- En la herramienta What If, haga clic en «Performance & Fairness» y, a continuación, seleccione «Ground Truth Feature = Over-50K» y «Slice by = GenderCode». Revise el conjunto de datos e inspeccione las variables clave. Preste especial atención a las características demográficas, como la raza/etnia y el género, y su posible correlación con los resultados de la preselección.

04 Analizar la equidad

Aplique métricas de equidad para analizar las diferencias en las tasas de precisión de rechazo entre los grupos demográficos, concretamente la raza/etnia.

Datapoint editor

Performance & Fairness

Features

600 datapoints loaded

Configure

Ground Truth Feature

Over-50K

Cost Ratio (FP/FN)

1

Slice by

RaceDesc

Slice by (secondary)

<none>

WHAT IS GROUND TRUTH?

The feature that your model is trying to predict. [More](#).

WHAT IS COST RATIO?

The cost of false positives relative to false negatives. Required for optimization. [More](#).

WHAT DOES SLICING DO?

Shows the model's performance on datapoints grouped by each value of the selected feature.

Fairness

Apply an optimization strategy

Select a strategy to automatically set classification thresholds, based on the set cost ratio and data slices. Manually altering thresholds or changing cost ratio will revert the strategy to 'custom thresholds'.

☒ Custom thresholds ⓘ

☐ Single threshold ⓘ

☐ Demographic parity ⓘ

☐ Equal opportunity ⓘ

☐ Equal accuracy ⓘ

☐ Group thresholds ⓘ

Custom thresholds for 5 values of RaceDesc ⓘ

Sort by

Count

Feature Value	Count	Model	Threshold ⓘ	False Positives (%)	False Negatives (%)	Accuracy (%)	F1	
▶ Hispanic	125	1	<div><div></div></div>	0.5 ⓘ	0.0	24.0	76.0	0.06
		2	<div><div></div></div>	0.5 ⓘ	3.2	20.8	76.0	0.29
▶ Other	124	1	<div><div></div></div>	0.5 ⓘ	0.0	28.2	71.8	0.00
		2	<div><div></div></div>	0.5 ⓘ	3.2	25.0	71.8	0.29
▶ Black	122	1	<div><div></div></div>	0.5 ⓘ	0.0	23.8	76.2	0.00
		2	<div><div></div></div>	0.5 ⓘ	4.9	18.9	76.2	0.41
▶ Asian	117	1	<div><div></div></div>	0.5 ⓘ	0.0	24.8	75.2	0.06
		2	<div><div></div></div>	0.5 ⓘ	5.1	19.7	75.2	0.47
▶ White	112	1	<div><div></div></div>	0.5 ⓘ	0.9	28.6	70.5	0.06
		2	<div><div></div></div>	0.5 ⓘ	8.0	21.4	70.5	0.44

05 Comparar resultados

Compare los resultados entre los distintos grupos, especialmente en lo que respecta al origen étnico. Determine si los candidatos con la misma cualificación pero de distintos orígenes étnicos reciben un trato desigual. Reflexione sobre cómo esto refleja los patrones conocidos de discriminación en la contratación.

06 Reflexionar sobre las implicaciones éticas

Debata si el modelo refleja o refuerza los prejuicios sociales y de qué manera. Evalúe la eficacia de las intervenciones que tienen en cuenta la equidad. Considere las implicaciones éticas más amplias del uso de la IA en la contratación y la gestión del talento.

• 05 Conclusión



Esta simulación ilustra cómo los modelos de aprendizaje automático, si no se controlan, pueden incorporar y reproducir sesgos históricos en los procesos de contratación automatizados. Aunque el conjunto de datos utilizado es sintético y el modelo se ha simplificado intencionadamente, los resultados reflejan las preocupaciones del mundo real en torno a la equidad en la contratación, especialmente en lo que respecta a las disparidades de género y étnicas.

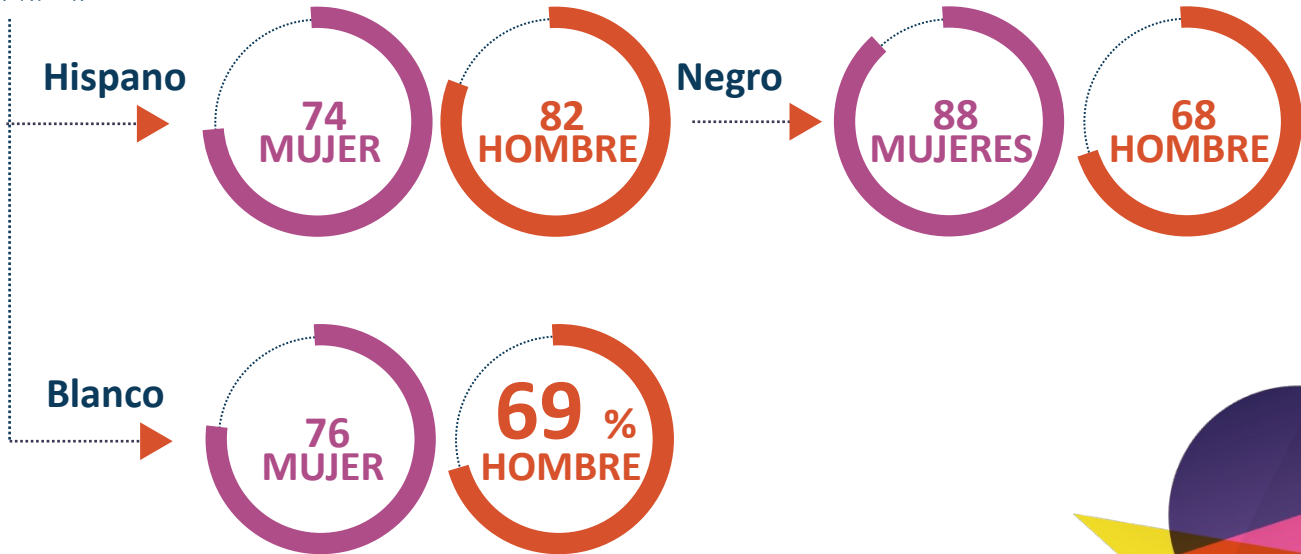
Los resultados revelan importantes inconsistencias en la precisión predictiva entre los distintos grupos demográficos. En concreto, la precisión de las predicciones de rechazo en la contratación es notablemente mayor en el caso de las mujeres

negras. En este contexto, una mayor precisión se corresponde con una menor tasa de falsos negativos, lo que significa que las solicitudes de las mujeres negras se clasifican con mayor frecuencia y corrección como rechazos.

Resumen de resultados



Precisión en la predicción de las decisiones de rechazo en la contratación, es decir, verdaderos «negativos» (%)



Esto sugiere que el modelo «recomienda» de manera desproporcionada rechazar a los candidatos de este grupo, en comparación con los de otros grupos, lo que pone de relieve un sesgo preocupante en el proceso de toma de decisiones del algoritmo.

Estos hallazgos reavivan los debates críticos sobre los puntos ciegos en la contratación algorítmica, en particular las formas en que las desigualdades estructurales se codifican en los datos de entrenamiento y posteriormente son aprendidas por los sistemas de IA. La observación de que la precisión predictiva varía entre los grupos marginados, e incluso dentro de las categorías protegidas, subraya la complejidad de evaluar la equidad: una mayor precisión no indica necesariamente resultados más equitativos.

Mediante la aplicación de métricas de equidad y el examen de los resultados a nivel de grupo, se invita a los estudiantes a comprometerse de forma crítica con las dimensiones sociotécnicas del sesgo algorítmico. Este ejercicio no solo proporciona experiencia práctica con técnicas de detección y mitigación del sesgo, sino que también hace hincapié en la responsabilidad ética de los desarrolladores y analistas en el diseño de sistemas de IA justos y responsables. En el contexto de la gestión de recursos humanos, la equidad no es solo una cuestión legal o de reputación, sino una obligación ética fundamental.

• 06 Referencias



- Allen, D. G., & Vardaman, J. M. (2017). Recruitment and Retention Across Cultures. *Annual Review of Organizational Psychology and Organizational Behavior*, 4(1), 153–181.
- Breugh, J. A. (2013). Employee Recruitment. *Annual Review of Psychology*, 64(1), 389–416.
- Czopp, A. M., Kay, A. C., & Cheryan, S. (2015). Positive Stereotypes Are Pervasive and Powerful. *Perspectives on Psychological Science*, 10(4), 451–463.
- Daly, M. C., Büchel, F., & Duncan, G. J. (2000). Premiums and penalties for surplus and deficit education: Evidence from the United States and Germany. *Economics of Education Review*, 19(2), 169–178.
- Ellemers, N. (2018). Gender Stereotypes. *Annual Review of Psychology*, 69(1), 275–298.
- HeModern Discrimination in Organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, M., Cheng, S. K., & Ng, L. C. (2020), 7(1), 257–282.
- Henderson, D. (2018). The Effects of Social Class on Perceptions of Job Applicants' Suitability for Employment. *Academy of Management Proceedings*, 2018(1), 13748.
- Hiemstra, A. M. F., Derous, E., Serlie, A. W., & Born, M. P. (2013). Ethnicity Effects in Graduates' Résumé Content. *Applied Psychology*, 62(3), 427–453.
- Horodyski, P. (2023). Recruiter's perception of artificial intelligence (AI)-based tools in recruitment. *Computers in Human Behavior Reports*, 10, 100298.
- Rigotti, C., & Fosch-Villaronga, E. (2024). Fairness, AI & recruitment. *Computer Law & Security Review*, 53, 105966.
- Seppälä, P., & Małecka, M. (2024). AI and discriminative decisions in recruitment: Challenging the core assumptions. *Big Data & Society*, 11(1), 20539517241235872
- Veit, S., & Thijsen, L. (2019). Almost identical but still treated differently: hiring discrimination against foreign-born and domestic-born minorities. *Journal of Ethnic and Migration Studies*, 1–20.
- Zaniboni, S., Kmicinska, M., Truxillo, D. M., Kahn, K., Paladino, M. P., & Fraccaroli, F. (2019). Will you still hire me when I am over 50? The effects of implicit and explicit age stereotyping on resume evaluations. *European Journal of Work and Organizational Psychology*, 28(4), 453–467.
- Zschirnt, E., & Ruedin, D. (2016). Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies*, 42(7), 1115–1134.



aileaders

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.