



www.aileaders-project.eu

Modelos sesgados de IA para la puntuación crediticia



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.

ESTUDIO DE CASO: 01

Este REA es un estudio de caso sobre cómo los datos sesgados o no representativos pueden dar lugar a resultados indeseables, discriminación y exclusión mediante el uso de servicios automatizados de puntuación crediticia. El estudio de caso analiza la aparición y el crecimiento de los sistemas automatizados de decisión crediticia que utilizan algoritmos para analizar diversos puntos de datos con el fin de evaluar la solvencia de una persona y decidir si se aprueba o se rechaza una solicitud de crédito.

El caso se centra en cómo funcionan los sistemas de puntuación crediticia, utilizando una simulación para ilustrar qué tipo de datos recopilan y qué tipo de datos personales utilizan para tomar decisiones. El estudio de caso también crea conciencia sobre los problemas que generan los conjuntos de datos sesgados o no representativos y cómo pueden hacer que los sistemas de puntuación crediticia sean injustos, funcionando como barreras y no como mejoras en la forma en que los ciudadanos pueden acceder a los servicios de crédito y préstamo en condiciones justas.



Objetivo/Finalidad

El objetivo de este estudio de caso es concienciar sobre el potencial de los sistemas automatizados de calificación crediticia para mejorar el acceso al crédito, así como sobre los retos y riesgos para la privacidad y la discriminación que conlleva el uso de estos sistemas, especialmente cuando se utilizan conjuntos de datos sesgados y también cuando los modelos se entrenan con datos que no son representativos de algunas poblaciones.



Resultados de aprendizaje esperados

- 01** El alumnado será capaz de **identificar** los riesgos éticos de los modelos de puntuación crediticia y proponer medidas correctivas.
- 02** El alumnado **comprenderá** las cuestiones éticas clave en las aplicaciones de IA para la concesión de préstamos y la puntuación crediticia, incluyendo el sesgo, la discriminación, la transparencia y la privacidad de los datos.

Enfoque metodológico sugerido

Este caso funciona mejor como aprendizaje basado en problemas, en el que el personal docente debe guiar un debate con el alumnado una vez que este se haya familiarizado con los conceptos relacionados con el acceso al crédito, los servicios de puntuación crediticia y las prácticas contemporáneas de solicitud de crédito y préstamo. Se proporcionan temas de debate, preocupaciones y posibles soluciones, pero el personal docente debe animar al alumnado a pensar por sí mismos e identificar otras posibles preocupaciones que puedan tener. Se proporcionan ejemplos y lecturas complementarias a través de enlaces, así como acceso a una simulación destinada a ilustrar qué tipo de datos personales se utilizan en la evaluación crediticia automatizada y qué tipo de resultados arrojan estos sistemas.

Palabras clave



Préstamos, puntuación crediticia, privacidad y protección de datos personales, transparencia, sesgo, discriminación.

Acceso automatizado al crédito

Es posible que haya visto varias ofertas en línea de compañías de tarjetas de crédito que prometen aprobar una tarjeta de crédito nueva (y la línea de crédito correspondiente) en tan solo uno o dos días. Algunas prometen calcular su límite de crédito en un par de minutos y autorizar su uso «de inmediato».

Es posible que haya visto varias ofertas en línea de compañías de tarjetas de crédito que prometen aprobar una tarjeta de crédito nueva (y la línea de crédito correspondiente) en tan solo uno o dos días. Algunas prometen calcular su límite de crédito en un par de minutos y autorizar su uso «de inmediato». Las solicitudes de crédito automatizadas en línea han facilitado y agilizado el proceso al calcular la puntuación de crédito de una persona casi al instante. Esto puede proporcionar a las empresas recomendaciones y/o aprobaciones automatizadas sobre una gama de productos financieros, no solo tarjetas de crédito, sino también otros productos como préstamos personales o hipotecas, al tiempo que calculan y hacen una oferta sobre los tipos de interés que se deben pagar por cada uno de estos productos.

Las puntuaciones crediticias son modelos informáticos que correlacionan una serie de factores con la probabilidad de que una persona incumpla el pago de su deuda. En otras palabras, una puntuación crediticia describe la solvencia de una persona o su capacidad y fiabilidad para devolver un préstamo determinado. Las puntuaciones crediticias se basan en el historial crediticio de una persona, construido en torno a una serie de datos, como el número de cuentas que tiene y el tiempo que lleva abierta cada una, el nivel total de deuda, el historial de pagos, los tipos de préstamos que tiene, la duración de sus transacciones crediticias, la proporción de deuda que está utilizando y si ha solicitado nuevas cuentas en un periodo de tiempo determinado. Mediante el uso de estos datos, las puntuaciones crediticias pueden separar los buenos riesgos crediticios de los malos y clasificar a los posibles prestatarios para predecir su probabilidad de impago.

En el sector financiero, los bancos y las instituciones financieras utilizan las puntuaciones crediticias para decidir sobre préstamos o tarjetas de crédito; las compañías de seguros las utilizan para evaluar el perfil de riesgo de los asegurados; los minoristas las utilizan para los planes de pago a plazos; los concesionarios de automóviles las utilizan para evaluar la elegibilidad y las condiciones de la financiación de automóviles; y los prestamistas hipotecarios las utilizan para evaluar la solvencia de los compradores de viviendas en el sector inmobiliario. Incluso las compañías telefónicas las utilizan para decidir si una persona es elegible para planes de prepago o pospago, o para determinar si pueden ofrecerle un teléfono nuevo financiado y pagadero a plazos. La combinación de múltiples puntos de datos y supuestos multivariantes se combina y pondera estadísticamente para permitir la automatización de las decisiones crediticias.

La inteligencia artificial y el aprendizaje automático

emplean modelos estadísticos y análisis de datos para agilizar y optimizar el proceso de evaluación crediticia. Esto permite una mayor rapidez y, potencialmente, una mayor precisión y confianza en las decisiones de préstamo. En los últimos años, tanto el sector bancario como el crediticio han comenzado a ofrecer varios productos **FinTech** (*tecnología financiera*) que ofrecen soluciones automatizadas para la toma de decisiones crediticias.

Las decisiones crediticias **basadas en la inteligencia artificial** pueden **mejorar** potencialmente **la eficiencia y el rendimiento, reduciendo los costes para las instituciones financieras, lo que también puede beneficiar al consumidor al ampliar el acceso al crédito o abaratar su coste**.¹ «Los nuevos modelos de alto rendimiento permiten a los bancos definir con mayor precisión los parámetros de concesión de préstamos (y de capital) y, por lo tanto, agudizar su capacidad para aprobar a los clientes solventes y rechazar las propuestas de los clientes que no son solventes o que no pueden permitirse más deuda. De hecho, los bancos (y las empresas de tecnología financiera) que han implantado estos nuevos modelos ya han aumentado sus ingresos, reducido las tasas de pérdida de crédito y obtenido importantes ganancias de eficiencia gracias a una toma de decisiones más precisa y automatizada².» Sin embargo, existen varios riesgos e inconvenientes que exploraremos durante este estudio de caso. Entre ellos se encuentra la falta de explicabilidad o la incapacidad de explicar por qué los programas toman determinadas decisiones. Otro es el impacto desigual o la discriminación que puede afectar a ciertos grupos más que a otros debido a las prácticas comerciales (como la forma en que se toman las decisiones de concesión de préstamos). Esto, como veremos, está directamente relacionado con el uso de datos defectuosos o sesgados para tomar decisiones crediticias, lo que puede dificultar el acceso al crédito a algunos colectivos, en lugar de facilitarlo, agravando así la desigualdad en el acceso al crédito³.

- 1 Servicio de Investigación del Congreso. (2023). *Inteligencia artificial generativa: visión general, cuestiones y preguntas para el Congreso* (Informe del CRS n.º IF12399). <https://crsreports.congress.gov/product/pdf/IF/IF12399>
- 2 Dash, R., Kremer, A. y Petrov, A. (2021). *Diseño de modelos de decisión crediticia de última generación*. McKinsey & Company. <https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/designing-next-generation-credit-decisioning-models>
- 3 Andrews, E. L. (6 de agosto de 2021). *Cómo los datos erróneos agravan la desigualdad en el crédito*. Stanford HAI. <https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit>

Modelos de datos y decisiones crediticias basadas en IA

Crédito basado en IA: ventajas específicas

Las decisiones crediticias tomadas con la ayuda de la inteligencia artificial tienen ventajas potenciales que van más allá de las mejoras en la eficiencia y la reducción de los costes asociados a la concesión de créditos a las personas, lo que los hace menos costosos.

Potencialmente, **el aprendizaje automático** en la puntuación crediticia también puede ampliar el número y la naturaleza de los puntos de datos que se tienen en cuenta para evaluar la solvencia de las personas, así como ofrecer la posibilidad de crear puntuaciones para comunidades que han estado crónicamente desatendidas por los bancos y conceder **acceso a soluciones bancarias y crediticias a comunidades enteras y a personas** que tradicionalmente han sido excluidas de las instituciones financieras formales. Al ofrecer alternativas mediante la incorporación de una gama más amplia de datos, **la puntuación basada en la inteligencia artificial** «permite evaluar a personas sin

historial crediticio tradicional mediante el examen de fuentes de datos alternativas, como transacciones en línea, interacciones en redes sociales, hábitos de navegación o uso de aplicaciones móviles».⁴ Para aquellos que cuentan con registros bancarios y financieros tradicionales, la puntuación basada en la IA puede proporcionar un análisis más preciso y mejorado para que las decisiones de puntuación crediticia y concesión de préstamos sean más exactas. Para aquellos que carecen de historiales tradicionales, puede ofrecer la posibilidad de acceder al crédito y a otros servicios bancarios.

Fuentes de datos tradicionales para la puntuación crediticia

- tarjetas de crédito y líneas de crédito abiertas
- préstamos para automóviles
- hipotecas
- historial de pagos de crédito
- historiales de consultas de crédito
- solicitudes de quiebra

Imagen 1. Ejemplos de datos crediticios tradicionales y alternativos.

Fuente: <https://www.afi-global.org/wp-content/uploads/2025/02/Alternative-Data-for-Credit-Scoring.pdf>

Fuentes alternativas de datos para la puntuación crediticia

- datos de flujo de caja
- pagos de facturas
- datos de alquiler
- registros de empleo
- registros de infracciones de tráfico o disputas
- patrones de uso de datos de telecomunicaciones y móviles
- datos de redes sociales
- datos de comportamiento
- pagos de impuestos

En la actualidad, una gran parte de la población mundial se considera consumidores invisibles (sin acceso) o con escaso acceso al crédito (acceso deficiente en malas condiciones). Solo en Estados Unidos, se estima que 45 millones de consumidores no reciben servicios o reciben servicios insuficientes por parte de los modelos de crédito tradicionales⁵. En países como India y Sudáfrica, más de la mitad de la población no tiene buenas opciones para acceder al crédito⁶. Allí donde había acceso, existe la posibilidad de hacerlo más justo y eficiente, mientras que allí donde no era posible acceder, existe la posibilidad de hacerlo finalmente accesible, al tiempo que se garantiza que este acceso sea igualmente eficiente y justo.

Existe un gran potencial para desarrollar modelos de calificación crediticia más inclusivos y sostenibles que, a su vez, puedan fomentar el progreso y la seguridad financiera de una población más amplia; más personas que nunca tienen la posibilidad de acceder a servicios bancarios y financieros que, a su vez, pueden dar lugar a productos crediticios más competitivos en el mercado, lo que puede acabar haciendo que el crédito sea más asequible⁷.

Entonces, ¿por qué nos preocupa que las decisiones crediticias basadas en la IA puedan aumentar, en lugar de disminuir, la exclusión financiera?

4 Bag, S. (2024). IA y puntuación crediticia: la ventaja algorítmica y las precauciones. Observer Research Foundation. <https://www.orfonline.org/expert-speak/ai-and-credit-scoring-the-algorithmic-advantage-and-precaution>

5 TransUnion. (2022). Más de 45 millones de estadounidenses no tienen acceso al crédito o tienen un acceso insuficiente; aproximadamente el 20 % pasa a tener crédito activo cada dos años. TransUnion. https://newsroom.transunion.com/more-than-45-million-americans-are-either-credit-unserved-or-underserved---approximately-20-mi-grate-to-being-credit-active-every-two-years#_edn1

6 Equipo de Svitla. (2024). Aprendizaje automático para la puntuación crediticia: ventajas, modelos y retos de implementación. Svitla Systems. <https://svitla.com/blog/machine-learning-for-credit-scoring/>

7 Alianza para la Inclusión Financiera. (2025). Datos alternativos para la calificación crediticia. <https://www.afi-global.org/wp-content/uploads/2025/02/Alternative-Data-for-Credit-Scoring.pdf>

Calificación crediticia y uso de datos personales

El proceso de aprobación de crédito: la puntuación crediticia y la aprobación de crédito implican procesos complejos que se desarrollan entre bastidores para que se pueda tomar una decisión.

Esto incluye datos proporcionados por el solicitante, el banco o la entidad financiera, una agencia de crédito y una agencia de puntuación crediticia. Una persona y las demás entidades proporcionan una gran cantidad de datos al proceso, entre los que se incluyen los datos personales. Aquí hay una imagen que muestra a grandes rasgos cómo funciona el proceso. Como se puede ver, no solo muestra las entidades involucradas, sino que también muestra las capas de seguridad utilizadas para proteger los datos empleados para la puntuación crediticia y la decisión posterior.

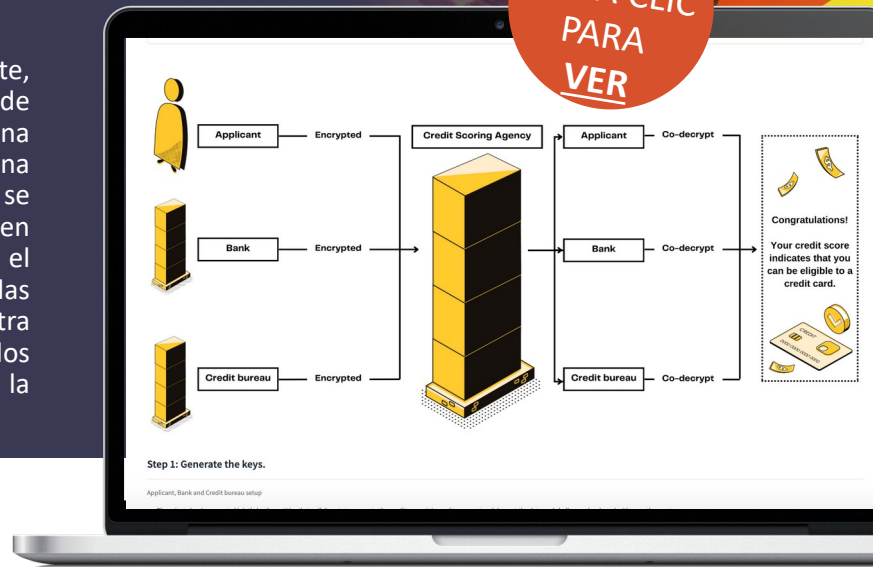


Imagen 2. Diagrama que describe el proceso de calificación crediticia y decisión de concesión de préstamos.

Fuente: https://huggingface.co/spaces/zama-fhe/encrypted_credit_scoring

El formulario muestra preguntas sobre el patrimonio y la situación personal del usuario. Las respuestas actuales son: 'Property' (seleccionado), '3' hijos, '3' miembros en el hogar, '35900' euros de ingresos anuales y '30' años de edad. Una etiqueta roja circular sobre el formulario dice 'HAGA CLIC PARA VER'.

Which of the following do you actively hold or own?

☐ Car ☒ Property ☐ Mobile phone

Number of children
How many children do you have ? 3

Household size
How many members does your household have ? 3

Income
What's your total yearly income (in euros) ? 35900

Age
How old are you ? 30

Para ilustrar cómo funciona este proceso y qué tipo de datos hay que introducir para que se evalúe la solvencia crediticia y se determine la concesión de acceso al crédito, como una tarjeta de crédito, **haga clic en la siguiente imagen**. Se mostrará una simulación de puntuación crediticia que proporcionará una evaluación de la probabilidad de que se apruebe una tarjeta de crédito.

Los retos y riesgos de datos sesgados y no representativos

Preocupaciones sobre la explicabilidad

Si bien los modelos de aprendizaje automático para la calificación crediticia y los préstamos ofrecen varias ventajas, también existen serias preocupaciones sobre la transparencia, la equidad y la posibilidad de que se refuercen los sesgos, lo que, a su vez, puede conducir a la exclusión crediticia, en lugar de a la inclusión.



Una preocupación específica es la **falta de explicabilidad** o la incapacidad de explicar por qué los programas de aprendizaje automático toman determinadas decisiones tras analizar los datos introducidos. Esto supone un problema para los usuarios de los sistemas, los reguladores y terceros, que pueden no ser capaces de comprender y explicar por qué un programa ha actuado de una determinada manera y por qué. La capacidad de la IA para «reaccionar ante grandes volúmenes de datos diversos, más allá del alcance de la capacidad cognitiva humana, es también **el talón de Aquiles del aprendizaje automático**, ya que dicha complejidad suele ser opaca en términos del proceso de toma de decisiones que precede a una decisión».⁸ El aprendizaje automático que toma decisiones tan complejas que no son fácilmente interpretables o explicables por los seres humanos se suele denominar «**aprendizaje automático de caja negra**» o «**IA de caja negra**».⁹

Esto supone un reto importante, ya que **limita la transparencia y la rendición de cuentas**. Cuando se deniega un préstamo a una persona solicitante, el prestamista debe poder explicar el motivo de la denegación. Si los prestamistas no pueden explicar suficientemente cómo los sistemas automatizados llegaron a la decisión, es posible que los solicitantes no dispongan de elementos suficientes para impugnar la decisión y se sientan indefensos,

lo que socava la confianza en la propia institución financiera. Además, los prestamistas deben asegurarse de que pueden justificar sus decisiones para cumplir con la legislación que puede exigir que las decisiones de crédito y préstamo tengan motivos claros que deben ser revelados a los consumidores.¹⁰ Otro problema es que **la IA de caja negra** obstaculiza los esfuerzos por mejorar el sistema. Si una decisión no es adecuada, «es **extremadamente difícil analizar por qué se ha cometido el error o determinar qué hay que hacer para corregir el modelo**».¹¹

Ser capaz de explicar las decisiones de la IA también es esencial para ganarse la confianza del usuario y garantizar que las decisiones sean justas y equitativas, por lo que **los métodos para auditar estos sistemas se vuelven esenciales**. La explicabilidad debe centrarse normalmente en explicar por qué «esta entrada concreta conduce a ese resultado concreto»¹², pero también es esencial saber qué datos internos conforman las estructuras de un programa concreto. En la siguiente sección, nos centramos en cuestiones relacionadas con los datos utilizados para entrenar los sistemas, que es lo que finalmente se combina con los datos concretos enviados por un usuario y permite al sistema de IA generar una decisión concreta.

8 King's College London. (s. f.). Los retos de la explicabilidad de la IA. <https://www.kcl.ac.uk/challenges-of-ai-explainability>

9 Kosinski, M. (29 de octubre de 2024). ¿Qué es la IA de caja negra y cómo funciona? IBM. <https://www.ibm.com/think/topics/black-box-ai>

10 Servicio de Investigación del Congreso. (2023). Inteligencia artificial generativa: visión general, cuestiones y preguntas para el Congreso (Informe del CRS n.º IF12399). <https://crsreports.congress.gov/product/pdf/IF/IF12399>

11 King's College London. (s. f.). Los retos de la explicabilidad de la IA. <https://www.kcl.ac.uk/challenges-of-ai-explainability>

12 Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M. y Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. ArXiv. <https://arxiv.org/abs/1806.00069>

La falta de diversidad de los conjuntos de datos y los datos defectuosos

Como se ha explicado anteriormente, los datos sesgados y los algoritmos sesgados pueden hacer que la toma de decisiones automatizada dé lugar a resultados que perjudiquen a quienes tienen dificultades para acceder a un buen crédito o incluso a los servicios financieros. Las minorías y los grupos de bajos ingresos suelen sufrir estos sesgos de manera desproporcionada.

Sin embargo, las investigaciones muestran que este no es el único problema. Los diferentes resultados para las minorías y las mayorías no solo están relacionados con el sesgo, sino también con el hecho de que los grupos minoritarios y de bajos ingresos tienen menos datos en sus historiales crediticios, ya que suelen estar infrarrepresentados en el acceso al crédito¹³. Esto significa que cuando estos datos se utilizan para calcular una puntuación crediticia y esta puntuación se utiliza para hacer una predicción sobre el impago de un préstamo, dicha predicción será menos precisa. Es esta falta de precisión la que conduce a la desigualdad, no solo el sesgo.¹⁴ La falta de diversidad en los conjuntos de datos para entrenar los modelos de aprendizaje automático daña a comunidades específicas, lo que aumenta la desigualdad. Las desigualdades sistémicas persisten en la curación

de conjuntos de datos y en la desigualdad de acceso. En algunos campos, también se debe a que hay desigualdad de oportunidades para participar en la creación de esos conjuntos de datos¹⁵. El rendimiento de cualquier sistema de IA está determinado en gran medida por los conjuntos de datos que analiza utilizando estadísticas, ya que sus resultados provienen de la identificación de patrones en los datos: «La calidad del conjunto de datos subyacente de un sistema de IA es crucial para su eficacia».¹⁶ En el contexto de los préstamos y el acceso al crédito, **«es un ciclo que se perpetúa a sí mismo... Concedemos préstamos a las personas equivocadas y una parte de la población nunca tiene la oportunidad de acumular los datos necesarios para que se le conceda un préstamo en el futuro»**.¹⁷

El riesgo del sesgo algorítmico

Si bien una de las ventajas del análisis automatizado de las solicitudes de crédito puede ser la reducción de la subjetividad en el proceso de toma de decisiones para la concesión de un préstamo, existe el riesgo de que estos procesos consoliden «los sesgos y prejuicios existentes contra grupos definidos por su raza, sexo, orientación sexual y otros atributos»¹⁸, algunos de los cuales son categorías especiales de datos personales protegidos por la ley.¹⁹

Esto se debe a que los conjuntos de datos suelen contener decisiones tomadas en el pasado por instituciones financieras, o porque no contienen datos suficientes sobre determinados grupos, lo que podría dar lugar a su discriminación. Si hace clic en la siguiente imagen, podrá acceder a un artículo de prensa que explica cómo el sistema

automatizado de aprobación de tarjetas de crédito de Apple Card, proporcionado por Goldman Sachs, dio lugar a quejas por presunta discriminación contra las solicitantes mujeres, que afirmaban que se les ofrecían límites de crédito más bajos o se les denegaba la tarjeta, incluso si sus maridos obtenían aprobaciones y mejores tasas.



- 13 Blattnr, L. y Nelson, S. (2021). ¿Cuánto cuesta el ruido? Datos y disparidades en el crédito al consumo. ArXiv. <https://arxiv.org/abs/2105.07554>
- 14 Heaven, W. D. (2021). El sesgo no es el único problema de las puntuaciones crediticias, y no, la IA no puede ayudar. MIT Technology Review. <https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/>
- 15 Arora, A., Alderman, J. E., Palmer, et al. (2023). El valor de las normas para los conjuntos de datos sanitarios en aplicaciones basadas en la inteligencia artificial. BMJ Health & Care Informatics, 30(1), e100888. <https://doi.org/10.1136/bmjhci-2023-100888>
- 16 Bach, A. y Kriebitz, A. (29 de septiembre de 2023). Diversidad en la IA: hacia una definición del problema. Human Technology Foundation. <https://www.human-technology-foundation.org/news/diversity-in-ai-towards-a-problem-statement>
- 17 Heaven, W. D. (2021). El sesgo no es el único problema de las puntuaciones crediticias, y no, la IA no puede ayudar. MIT Technology Review. <https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/>
- 18 García, A. C. B., García, M. G. P. y Rigobon, R. (2024). Discriminación algorítmica en el ámbito crediticio: ¿qué sabemos al respecto? AI & Society, 39, 2059-2098. <https://doi.org/10.1007/s00146-023-01676-3>
- 19 Artículo 9 del Reglamento General de Protección de Datos: <https://gdpr-info.eu/art-9-gdpr/>

La **denegación del acceso a los mercados crediticios puede ser discriminatoria** cuando las decisiones se basan en sesgos preexistentes. También puede manifestarse debido a un trato diferencial que ofrece condiciones más desfavorables a quienes sufren discriminación, como comisiones diferentes y menos ventajosas o tipos de interés más elevados.²⁰

Los sesgos en los conjuntos de datos pueden **erigir nuevas barreras para el acceso a los servicios financieros de las poblaciones tradicionalmente desatendidas** que ya tenían dificultades para acceder a estos servicios. Los sistemas tradicionales de puntuación crediticia penalizan a quienes no tienen un historial crediticio formal, pero incluso los sistemas que utilizan datos no tradicionales, si muestran sesgos en sus decisiones o refuerzan los sesgos de sus creadores, pueden **reforzar la exclusión**, lo que afecta de manera desigual a esas

comunidades desatendidas.

El acceso limitado a los servicios financieros tradicionales por parte de determinados colectivos puede perpetuar las disparidades de riqueza. La discriminación pasada en este ámbito ha tenido efectos a largo plazo en la riqueza generacional global de determinados grupos, con consecuencias como «una propiedad inmobiliaria limitada, menores oportunidades empresariales y diferencias de riqueza generacionales».²¹

Si no se diseñan y aplican con cuidado, **los modelos de IA «perpetuarán las desigualdades»**²², anulando cualquier posible impacto positivo de un proceso de toma de decisiones que reduce la subjetividad solo sobre el papel, ya que conlleva los sesgos estructurales ya presentes en la sociedad.

¿De dónde provienen los sesgos?

Los sesgos algorítmicos pueden provenir de diferentes fuentes, **incluidos los datos de entrenamiento sesgados**, que pueden ser datos históricos que reflejan los sesgos sociales existentes. Esto se conoce como «sesgo de entrada, sesgo de salida» o cuando un modelo se entrena con datos relacionados con resultados ya sesgados. Es probable que la IA simplemente replique los problemas del pasado.

La **falta de diversidad** no solo se refleja en los datos. También puede aparecer en **los equipos de diseño de producto**: «La homogeneidad entre los equipos de ciencia de datos y de desarrollo contribuye a perpetuar los sesgos en los sistemas de IA».²³ La falta de experiencia o de comprensión de las comunidades desfavorecidas puede hacer que quienes diseñan los sistemas y proporcionan los datos de entrenamiento sean ajenos a sus dificultades y retos o a sus situaciones específicas, lo que puede dar lugar a un análisis inexacto de sus historiales crediticios.

Las decisiones sesgadas **pueden reforzarse y afianzarse con el tiempo**, a medida que los sistemas realizan evaluaciones cada vez más incorrectas que luego se consideran válidas, generando **bucles de retroalimentación** en los que las denegaciones incorrectas de crédito para un colectivo específico se reflejan en **los datos de entrenamiento futuros, perpetuando aún más la desigualdad y la injusticia.**²⁴

20 García, A. C. B., García, M. G. P., & Rigoban, R. (2024). Discriminación algorítmica en el ámbito crediticio: ¿qué sabemos al respecto? *AI & Society*, 39, 2059-2098. <https://doi.org/10.1007/s00146-023-01676-3>

21 Nuka, T. F. y Osedahunsi, B. O. (2024). Del sesgo al equilibrio: integración de la DEI en los sistemas financieros impulsados por la IA para promover la equidad crediticia. *International Journal of Science and Research Archive*, 13(2), 1189-1206. <https://doi.org/10.30574/ijrsra.2024.13.2.2257>

22 Nuka, T. F. y Osedahunsi, B. O. (2024). Del sesgo al equilibrio: integración de la DEI en los sistemas financieros impulsados por la IA para promover la equidad crediticia. *Archivo de la Revista Internacional de Ciencia e Investigación*, 13(2), 1189-1206. <https://doi.org/10.30574/ijrsra.2024.13.2.2257>

23 Nuka, T. F. y Osedahunsi, B. O. (2024). Del sesgo al equilibrio: integración de la DEI en los sistemas financieros impulsados por la IA para promover la equidad crediticia. *Archivo de la Revista Internacional de Ciencia e Investigación*, 13(2), 1189-1206. <https://doi.org/10.30574/ijrsra.2024.13.2.2257>

24 Nuka, T. F. y Osedahunsi, B. O. (2024). Del sesgo al equilibrio: integración de la DEI en los sistemas financieros impulsados por la IA para promover la equidad crediticia. *Archivo de la Revista Internacional de Ciencia e Investigación*, 13(2), 1189-1206. <https://doi.org/10.30574/ijrsra.2024.13.2.2257>

Directrices para el personal docente

Sobre el estudio de caso

Este estudio de caso ofrece al personal docente la posibilidad de fomentar el debate en clase sobre varios temas relacionados con la **inclusión y la exclusión financieras**, entre ellos las **posibilidades y los riesgos de la calificación crediticia automatizada y las decisiones crediticias automatizadas**.

En primer lugar, podría comenzar con un **debate sobre la puntuación crediticia basada en la inteligencia artificial y los servicios financieros basados en la inteligencia artificial**, discutiendo las **ventajas de la automatización** y el **potencial del uso de diferentes tipos de conjuntos de datos para prestar servicio a diversas poblaciones y centrarse en la inclusión**.

Después de aprender sobre el **proceso de aprobación de créditos**, debe **centrarse en el uso de datos personales para este tipo de decisiones**, ya que los modelos basados en IA requieren conjuntos de datos no solo para el entrenamiento, sino también para que los usuarios introduzcan datos personales, como se muestra en la **simulación de la página 8 de este estudio de caso**.

Aquí hay algunas preguntas para su alumnado:



- ¿Creéis que existen **riesgos para la privacidad y la protección de los datos personales de las personas?**
- Si es así, ¿por qué crees que **es importante proteger los datos personales?**
- ¿Qué **consecuencias** podrían sufrir los clientes si sus **datos se hicieran públicos?**
- Además de las fugas de datos y otras amenazas a la ciberseguridad, ¿qué **otros riesgos pueden asociarse al uso de datos personales para tomar decisiones sobre la concesión de préstamos?**

Una vez que llegue a la cuarta parte, **debe debatir los retos con el alumnado**, uno por uno y, a **continuación, centrarse en las soluciones**. El estudio de caso incluye algunas referencias y lecturas adicionales en los anexos que figuran a continuación, pero un ejercicio interesante consiste en dividir al alumnado en grupos y pedirles que investiguen para **encontrar posibles soluciones a cada uno de los problemas planteados**. Uno o varios grupos pueden centrarse en la investigación sobre cuestiones de explicabilidad, otros en soluciones relacionadas con el problema de la diversidad en los conjuntos de datos y otros en los sesgos algorítmicos en general. Después de investigar, puede pedir al alumnado que **haga una breve presentación para exponer las soluciones que podrían implementar**.

Puede indicar al alumnado que **busque soluciones que abarquen desde métodos normativos y legislativos hasta soluciones más técnicas**. Existen excelentes recursos en línea que tratan estos temas, por lo que también es **una buena oportunidad para**

que realicen una investigación en la que piensen de forma crítica e identifiquen fuentes fiables que propongan soluciones válidas.

Por último, si cree que encaja en el contexto de su clase, puede **debatir las ventajas y los inconvenientes de la tendencia general a automatizar la toma de decisiones que pueden afectar a la vida de las personas, así como la necesidad de que, al menos, haya seres humanos que supervisen estas decisiones**. Puede **plantear un debate interesante sobre la necesidad de la interacción humana y la concesión de oportunidades a los solicitantes de crédito frente a las ventajas de las decisiones asistidas por la automatización o incluso delegadas a sistemas de IA y el potencial de las decisiones que eliminan la subjetividad**. ¿Es esto deseable? ¿Es siquiera posible? O, por el contrario, ¿es necesario cierto grado de subjetividad y agencia humana para tomar decisiones justas?



Lecturas adicionales

- Alliance for Financial Inclusion. (2025). Alternative data for credit scoring. <https://www.afi-global.org/wp-content/uploads/2025/02/Alternative-Data-for-Credit-Scoring.pdf>
- Alonso, A., & Carbó, J. M. (2022). Accuracy of explanations of machine learning models for credit decisions (Working Paper No. 2222). Banco de España. <https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/22/Files/dt2222e.pdf>
- Blattner, L., & Nelson, S. (2022). How flawed data aggravates inequality in credit. Stanford Graduate School of Business. <https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit>
- Dash, R., Kremer, A., & Petrov, A. (2021). Designing next-generation credit-decisioning models. McKinsey & Company. <https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/designing-next-generation-credit-decisioning-models>
- Heaven, W. D. (2021). Bias isn't the only problem with credit scores – and no, AI can't help. MIT Technology Review. <https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/>
- Kelion, L. (2019). Apple's 'sexist' credit card investigated by US regulator. BBC News. <https://www.bbc.com/news/business-50365609>
- Mensalvas, E., Guzmán, M. A. & Ruiz Bonilla, S. (2022). ML applied to credit risk: Building explainable models. <https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2022/10/Idanae-3Q22.pdf>
- ODSC. (2022). Should AI decide who gets a loan? Medium. <https://odsc.medium.com/should-ai-decide-who-gets-a-loan-83c6f259081b>

