



PROPUESTA TRABAJO FIN DE GRADO

Sesgos de la Inteligencia Artificial en el sector financiero

Paula López de Uralde Lozano

E2+BBAA

ÍNDICE

1.	TÍTULO.....	2
2.	RESUMEN.....	2
3.	OBJETIVOS.....	3
4.	JUSTIFICACIÓN	3
4.1.	ACADÉMICA.....	3
4.2.	SOCIAL Y PROFESIONAL.....	4
4.3.	PERSONAL.....	4
5.	METODOLOGÍA	5
6.	BIBLIOGRAFÍA.....	6

1. TÍTULO

<<Sesgos de la Inteligencia Artificial en el sector financiero>>

2. RESUMEN

El propósito de este trabajo será investigar los diferentes sesgos que presenta la Inteligencia Artificial en el sector financiero y como estos pueden afectar a ciertos colectivos.

Se estudiará como la Inteligencia Artificial se ha convertido en una herramienta esencial en el sector y ha sustituido a los métodos tradicionales para la toma de decisiones en procesos financieros como la concesión de créditos y evaluación de riesgos. Esto nace como consecuencia de su capacidad para procesar en tiempo real una gran cantidad de datos, lo que se traduce en una optimización del tiempo y de los recursos de las empresas.

Si nos centramos en el sector financiero, la gestión de grandes volúmenes de datos supone una parte muy importante ya que permite tomar decisiones más informadas y precisas en un entorno que está en constante crecimiento. Esta escalabilidad frente a la que proporcionan los métodos tradicionales permite manejar una cantidad exponencialmente mayor de datos. Además, permite la identificación de patrones y la predicción de tendencias del mercado lo que facilita la gestión de inversiones y la reducción de los riesgos asociados a estas.

Sin embargo, el uso de esta tecnología ha puesto de manifiesto la presencia de sesgos que pueden influir negativamente en determinados grupos sociales, aumentando las barreras para acceder a financiación y profundizando las desigualdades económicas y sociales. Esto se ha demostrado en estudios recientes, como el realizado por la Comisión Europea en su "Ethics Guidelines for Trustworthy AI", donde se destaca que el sesgo en algoritmos de IA puede generar exclusión financiera para minorías y otros colectivos vulnerables. Además, investigaciones de organizaciones como el MIT y el Berkeley Artificial Intelligence Research Lab han documentado cómo las IA entrenadas con datos históricos tienden a replicar y perpetuar sesgos discriminatorios presentes en dichos datos.

Para crear un sistema de Inteligencia Artificial que funcione correctamente, es necesario un Data Lake que contenga un conjunto de datos históricos suficientemente grande y representativo para poder generar y entrenar el algoritmo. Estos datos históricos pueden contener algún patrón que refuerce estos sesgos y que resulten en decisiones financieras que afecten negativamente a estos colectivos e intensifique estas diferencias. Según un estudio de ProPublica en 2016, los algoritmos utilizados en la evaluación de riesgos crediticios mostraron un sesgo sistemático contra personas de color, subrayando cómo los datos históricos reflejan disparidades sociales)

En el trabajo se abordarán los tipos de sesgos que aparecen en los sistemas de IA utilizados en el sector financiero, así como su impacto en ciertos colectivos. También se plantearán una serie de medidas para mitigar estos sesgos promoviendo la transparencia y la equidad, evitando así la desigualdad en el acceso a financiación de determinados colectivos. Tal y como indica ChatGPT sobre algunas iniciativas que están en marcha, existen medidas que han sido discutidas por expertos en ética de la IA, como la creación de "algoritmos explicables" o "auditorías algorítmicas", recomendadas por el IEEE y la UNESCO para mejorar la transparencia y minimizar el impacto de los sesgos. De esta forma se busca evitar la desigualdad en el acceso a la financiación y reducir las barreras que enfrentan ciertos grupos.

3. OBJETIVOS

(PONER LOS OBJETIVOS EN VERBO) El objetivo principal es estudiar los sesgos de la Inteligencia Artificial en el sector financiero y analizar cómo estos afectan a ciertos colectivos. También se propondrán estrategias para mitigar estos sesgos basándonos en los hallazgos del análisis que llevemos a cabo. Para lograr esto, se llevará a cabo un estudio de los distintos tipos de sesgos que presenta la Inteligencia Artificial en las finanzas y su clasificación, observando cómo cada tipo de sesgo influye en decisiones financieras.

Se identificarán y clasificarán los tipos de sesgos que aparecen en la aplicación de la Inteligencia Artificial en las finanzas. La naturaleza de estos sesgos puede ser diversa, y el estudio y clasificación de estos facilitará la interpretación del problema para el análisis que se llevará a cabo.

Tras el análisis y clasificación de los sesgos, se observará el impacto que tiene en ciertos colectivos como minorías étnicas o personas con bajos ingresos. Se analizarán situaciones en las que la aplicación de la Inteligencia Artificial presenta sesgos como pueden ser sistemas de detección de fraude, concesiones crediticias, acceso a financiación y algoritmos de scoring crediticio.

El estudio demostrará la presencia de sesgos y explicará las consecuencias que estos conllevan para los colectivos afectados. Además, estas decisiones intensifican las desigualdades económicas entre los distintos grupos sociales, reduciendo y limitando su capacidad financiera. Este impacto negativo subrayará la importancia de analizar este problema desde ambas perspectivas, social y económica.

Basándonos a las conclusiones del análisis, se desarrollarán una serie de estrategias y recomendaciones que ayuden a eliminar estos sesgos. Estas estrategias y recomendaciones pueden enfocarse en mejorar la calidad de los datos evitando datos sesgados para que a la hora de entrenar los algoritmos sean más transparentes y por tanto más fáciles de interpretar.

Finalmente, con este trabajo se busca demostrar que existen una serie de sesgos ligados a la aplicación de la Inteligencia Artificial en las finanzas, que a su vez promueven desigualdades en la sociedad intensificando diferencias sociales. Con esta investigación, se propondrán una serie de recomendaciones que mitiguen estos sesgos.

4. JUSTIFICACIÓN

4.1. ACADÉMICA

Es necesario abordar este tema ya que existe una creciente preocupación por los sesgos de la Inteligencia Artificial pero los estudios sobre su impacto en el sector financiero y en ciertos colectivos sociales son limitados. A pesar de ser un tema de interés actual dado que su uso está sustituyendo a los métodos tradicionales, los estudios se centran en aplicaciones generales en distintos sectores y en aspectos técnicos enfocados en mejorar su rendimiento y precisión en las finanzas, pero sin profundizar en el impacto que tienen los sesgos en distintos colectivos.

Para demostrar esta afirmación, he realizado una búsqueda en Google Scholar ("ML Bias in Fin Services") y he encontrado que la mayoría de los estudios se centran en mejorar la precisión y eficiencia de la Inteligencia Artificial dejando en un segundo plano el análisis del impacto social y ético. Es decir, la mayoría de las investigaciones están centradas en optimizar el rendimiento técnico de los modelos de Machine Learning aplicados en el sector financiero sin ir más allá en las implicaciones que los sesgos presentes pueden generar en colectivos vulnerables.

Un claro ejemplo que refuerza esta afirmación es el estudio "Machine Learning and AI in Financial Services: Market Developments and Financial Stability Implications" del Financial Stability Board (2017) ya que analiza todas las aplicaciones de la IA y el Machine Learning en las finanzas, evaluando como estas tecnologías mejoran la eficiencia de los mercados financieros. Sin embargo, deja a un lado los impactos que tienen los sesgos que aparecen tras la aplicación de esta tecnología.

Este estudio es relevante, actual y pertinente ya que además de haber sido realizado por una institución como el Financial Stability Board, menciona una tendencia que se está dando en la literatura sobre la IA en las finanzas que consiste en enfocarse en los beneficios técnicos de la automatización y optimización de procesos, mientras que las preocupaciones sobre la equidad, la transparencia y los impactos sociales quedan en segundo plano. Al no mencionar estos sesgos, se omite un aspecto muy importante en el uso de estas tecnologías, como es el riesgo de amplificar e intensificar las desigualdades sociales en el acceso al crédito.

La importancia de este trabajo no solo aparece en el uso de la IA como una herramienta técnica sino como una tecnología con un impacto social significativo, cobrando así un enfoque más innovador.

4.2. SOCIAL Y PROFESIONAL

En lo que se refiere a la perspectiva social, la mitigación de los sesgos de la Inteligencia Artificial en el sector financiero puede tener un impacto positivo, mejorando la equidad y eliminando las diferencias sociales a la hora del acceso a financiación de ciertos colectivos. Esto se alinea directamente con el *Objetivo de Desarrollo Sostenible 10* (ODS 10), cuyo propósito es disminuir la desigualdad. Eliminar los sesgos algorítmicos no solo favorece una mayor inclusión financiera, sino que también ayuda a garantizar que colectivos marginados, como mujeres, minorías étnicas o personas con bajos ingresos, no enfrenten barreras adicionales a la hora de solicitar créditos u otros servicios financieros.

Desde un punto de vista profesional, las empresas pueden mejorar la satisfacción de los clientes demostrando que sus decisiones financieras son justas y los algoritmos que se emplean son transparentes. Además, un sistema financiero justo permite abrir nuevas oportunidades de mercado atendiendo a una base de clientes más grande y diversa. Con todo ello, además de tratarse de una cuestión ética, la implementación de prácticas que eliminen estos sesgos se puede traducir en ventajas competitivas y beneficios económicos para las empresas.

4.3. PERSONAL

A nivel personal, este tema ha despertado mi interés ya que la Inteligencia Artificial es un tema actual y sus aplicaciones son cada vez más comunes en distintos sectores. La rapidez con la que la IA ha evolucionado me ha despertado mucha curiosidad, no solo en sus implicaciones a nivel técnico, pero

sobre todo a cómo afectará en lo social y lo ético. Esto me ha hecho plantearme como puede influir en el mundo financiero y que riesgos puede acarrear si no es gestionado de una forma correcta.

Por otro lado, mi interés académico por las finanzas hace que estudiar la aplicación de la Inteligencia Artificial en este campo me resulte algo muy interesante y útil de cara al futuro.

Los avances en la Inteligencia Artificial los últimos meses me hacen pensar la importancia que ha cobrado rápidamente y como esto puede conllevar riesgos futuros si no es gestionada de forma adecuada. La investigación de este tema puede tener un impacto positivo tanto a nivel social como a nivel profesional y me parece muy interesante la idea de combinar una investigación que pueda tener un impacto social y contribuir a un cambio positivo para la sociedad.

5. METODOLOGÍA

Para el estudio de los sesgos de la Inteligencia Artificial en el sector financiero y su impacto en colectivos específicos, se empleará una metodología cualitativa, aunque también aparecerán elementos cuantitativos ya que ayudará a comprender mejor el impacto de estos sesgos y permitirán un análisis más completo. Esta combinación de ambas metodologías permitirá examinar tanto las causas de los sesgos como el impacto que estos tienen en ciertos colectivos.

La investigación empezará con un análisis de los estudios existentes sobre los sesgos de la Inteligencia Artificial tanto en el sector financiero como en otros sectores (sanitario, educación, RRHH, seguridad...). Este análisis permitirá concretar las bases del estudio y a partir de ahí se identificarán las posibles preguntas que podrán guiar la investigación hacia un enfoque más preciso.

También se analizarán casos específicos donde se hayan identificado sesgos en sistemas de Inteligencia Artificial en finanzas. Estos casos pueden tratarse tanto de concesiones de créditos, evaluación de riesgos o decisiones de inversión. En este análisis se empleará una metodología cuantitativa ya que permitirá una comprensión detallada de estos sesgos y de sus impactos.

Finalmente se desarrollarán estrategias para mitigar los sesgos en la Inteligencia Artificial en finanzas, basadas en los hallazgos de la investigación previa respaldadas con el análisis cuantitativo. Se tendrán en cuenta factores como el entorno financiero actual, la normativa actual y el impacto que tendrían en la transparencia del sector financiero.

En resumen, esta combinación de la metodología cuantitativa y cualitativa permitirá un análisis detallado de los sesgos de la Inteligencia Artificial en el sector financiero partiendo de una base cualitativa y apoyado por datos cuantitativos para una mejor comprensión del tema proporcionando soluciones efectivas al problema.

6. BIBLIOGRAFÍA

1. Getronics. *AI and the finance industry: A perfect match*.
<https://www.getronics.com/es/ai-and-the-finance-industry-a-perfect-match/#:~:text=La%20IA%20se%20ha%20convertido,detectar%20y%20prevenir%20actividades%20fraudulentas>.
2. Nexus Integra. Ventajas y desventajas de la inteligencia artificial.
<https://nexusintegra.io/es/ventajas-y-desventajas-de-la-inteligencia-artificial/#:~:text=La%20Inteligencia%20artificial%20permite%20que,autom%C3%A1tica%20y%20sin%20intervenci%C3%B3n%20humana.&text=La%20IA%20libera%20a%20las,tiempo%20a%20desarrollar%20funciones%20creativas>.
3. Silicon. Uso de la inteligencia artificial en la concesión de créditos.
<https://www.silicon.es/uso-inteligencia-artificial-creditos-2420797>
4. 1MillionBot. ¿Cómo se entrena un modelo de IA?
<https://1millionbot.com/como-se-entrena-un-modelo-de-ia/>
5. Hewlett Packard Enterprise. ¿Qué es un data lakehouse?
<https://www.hpe.com/es/es/what-is/data-lakehouse.html>
6. NTT Data. IA en el sector financiero: Ejemplos reales de aplicación.
<https://co.nttdata.com/insights/blog/ia-en-el-sector-financiero-ejemplos-reales-de-aplicacion>
7. Hachaichi, Y. Artificial intelligence and machine learning in financial services: Market developments and financial stability implications.
https://www.researchgate.net/profile/Yassine-Hachaichi/publication/369978046_Artificial_intelligence_and_machine_learning_in_financial_services_Market_developments_and_financial_stability_implications/links/6437d64e4e83cd0e2facd021/Artificial-intelligence-and-machine-learning-in-financial-services-Market-developments-and-financial-stability-implications.pdf
8. Cannarsa, M. (2021). Ethics guidelines for trustworthy AI. The Cambridge Handbook of Lawyering in the Digital Age.
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
9. Angwin, J., Larson, J., Kirchner, L., & Mattu, S. (2016, mayo 23). Machine bias.
<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>