



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

Doble Grado E2 + Business Analytics

TRABAJO DE FIN DE GRADO

Impacto de la Inteligencia Artificial en la productividad y competitividad empresarial: un análisis bibliométrico con foco en banca de inversión

Autor: Álvaro Olivares Gómez de Barreda

Director: Víctor Pérez Segura

Curso 2025–2026

INDICE

1	Introducción:	
1.1	Contexto y motivación.....	3
1.2	Justificación y relevancia.....	4
1.3	Pregunta general y objetivo general.....	4
1.4	Estructura del trabajo.....	4
2	Marco teórico:	
2.1	Productividad y competitividad empresarial en la era digital.....	5
2.2	Aplicaciones de la IA en banca de inversión.....	6
2.3	Retos y resistencias a la adopción.....	7
3	Metodología	
3.1	Justificación del enfoque bibliométrico.....	8
3.2	Preguntas y objetivos específicos.....	9
3.3	Fuente de datos: Scopus.....	11
3.4	Diseño de la query de búsqueda: proceso iterativo de prueba y error.....	11
3.5	Proceso de filtrado y diagrama PRISMA.....	15
3.6	Variables exportadas.....	17
3.7	Herramientas de análisis.....	17
4	Análisis de datos y resultados	
4.1	Evolución temporal del estudio sobre la IA y banca de inversión.....	18
4.2	Autores y países más representativos dentro del ámbito (PE2).....	19
4.3	Temas principales: análisis de palabras clave (PE4).....	21
4.4	Posibles tendencias temáticas emergentes (PE5).....	23
4.5	Topic modelling (PE4 + PE5).....	24
4.6	Revistas con mayor producción.....	27
5	Conclusiones, limitaciones y futuras líneas de investigación	
5.1	Discusión de resultados obtenidos.....	29
5.2	Limitaciones del trabajo.....	31
5.3	Futuras líneas de investigación.....	31
5.4	Reflexión final.....	32
6	Bibliografía.....	33
7	Anexo: código de Python utilizado en la sección 4.....	34

1. Introducción

1.1. Contexto y motivación

La inteligencia artificial (IA) se ha consolidado en los últimos años como una de las tecnologías con mayor capacidad transformadora sobre los modelos de negocio y los procesos de trabajo en el mundo laboral. Su aplicación en sectores como el comercio, la salud, la fabricación o el marketing ha generado ganancias en eficiencia operativa, calidad analítica y capacidad de trabajo. Sin embargo, la adopción de esta tecnología disruptiva no ha sido homogénea entre industrias: en sectores altamente regulados, donde se maneja información sensible y con una fuerte cultura de evitar riesgo reputacional, la incorporación de la IA avanza con mayor lentitud. La banca de inversión es un caso claro de esta dinámica.

En banca de inversión se opera con grandes volúmenes de datos estructurados y no estructurados. La mayoría de las tareas diarias son analíticas, complejas y repetitivas (modelos financieros en Excel, documentos de due diligence, análisis de empresas cotizadas, presentación de datos y estrategias en PowerPoint, valoración de activos), en un entorno donde la rapidez y la calidad del análisis son determinantes. A priori, cumple con las características de una industria ideal para la automatización y para los sistemas de ayuda a la decisión basados en IA. No obstante, la adopción de estas tecnologías en el día a día del sector ha sido más lenta que en otras áreas de los servicios financieros, como la banca minorista o la gestión algorítmica de carteras.

La motivación personal de este trabajo nace de mi experiencia profesional. Durante mi periodo de prácticas en banca de inversión el verano pasado, pude observar la enorme cantidad de tareas analíticas y operativas que podrían ser optimizadas mediante herramientas de IA, así como las resistencias culturales, organizativas y regulatorias dentro de los bancos que ralentizan su integración. Esa observación, junto con la velocidad con la que la IA está transformando otros sectores, genera en mí un interés por comprender en qué punto se encuentra la investigación académica sobre el tema y qué dirección está tomando.

El presente Trabajo Fin de Grado (TFG) aborda esta cuestión desde una perspectiva bibliométrica. En lugar de estudiar un caso concreto o realizar entrevistas a profesionales, se propone un análisis sistemático y cuantitativo de la literatura científica internacional publicada sobre IA, productividad y competitividad empresarial, prestando especial atención al ámbito de la banca de inversión. El análisis bibliométrico permite centrarse en un entorno adecuado

de investigación, identificar quién está liderando la investigación, detectar las modas y opiniones dominantes, y revelar los vacíos existentes o posibles escenarios en un futuro respecto al tema de investigación que se propone.

1.2. Justificación y relevancia

El estudio presentado a continuación es relevante por tres elementos. En primer lugar, la IA constituye una de las tecnologías más disruptivas y con impacto directo sobre la productividad empresarial en la próxima década, lo que justifica el interés por entender qué se piensa a nivel académico sobre el tema y que implicaciones puede tener en el sector al que me dirijo yo al principio de mi carrera profesional. En segundo lugar, la banca de inversión es un sector económicamente significativo cuya transformación digital podría tener implicaciones sistémicas, pero al mismo tiempo se encuentra menos representado en la literatura académica sobre IA aplicada que otros segmentos financieros. Existe, por tanto, un vacío de conocimiento que puede ser relevante de estudiar. En tercer lugar, el enfoque bibliométrico aporta una perspectiva sistemática que puede ofrecer una visión de conjunto complicada de encontrar mediante estudios de investigación puntuales o tradicionales.

1.3. Pregunta general y objetivo general

El trabajo gira en torno a una pregunta general que orienta toda la investigación:

¿Cómo ha evolucionado la investigación científica internacional sobre el impacto de la inteligencia artificial en la productividad y competitividad en el mundo laboral, y sobre todo, en la banca de inversión, y qué tendencias presenta para el futuro?

De esta pregunta se deriva el objetivo general del trabajo: analizar la evolución y tendencias de la literatura científica internacional sobre el impacto de la IA en la productividad y competitividad de la banca de inversión, mediante técnicas bibliométricas aplicadas a un corpus de artículos indexados en Scopus. A medida que la investigación avanza, se irán formulando una serie de preguntas más específicas y de objetivos específicos, que se presentan en el capítulo de metodología, una vez llevadas a cabo las técnicas analíticas necesarias.

1.4. Estructura del trabajo

La investigación se divide en cinco capítulos. Tras esta introducción, el capítulo 2 desarrollará el marco teórico, abordando los conceptos de productividad y competitividad empresarial en

la era digital, las aplicaciones concretas de la IA en banca de inversión y los principales obstáculos a su adopción. En el capítulo 3 detallaré la metodología bibliométrica utilizada, explicando el proceso iterativo de construcción de la base de datos desde Scopus, el diseño de las queries de búsqueda, los criterios de inclusión y exclusión, las variables consideradas y el diagrama PRISMA del proceso de filtrado manual. Una vez creada la base de datos y tras haber analizado los artículos en Python, en el capítulo 4 presentaré los resultados del análisis bibliométrico. Por último, el capítulo 5 recogerá la discusión y las conclusiones del trabajo, junto con sus limitaciones y posibles líneas futuras de investigación.

2. Marco teórico

2.1. Productividad y competitividad empresarial en la era digital

Los conceptos de productividad y competitividad son claves en el mundo laboral a nivel económico y de gestión de negocios. La productividad, entendida como la relación entre los recursos utilizados y los resultados obtenidos, es uno de los principales factores de crecimiento económico y operativo, sirviendo como métrica clave para valorar el rendimiento de empresas. La competitividad, por su parte, hace referencia a la capacidad de una empresa para mantener o ampliar su posición en un mercado frente a sus rivales, y depende de factores que van desde la estructura de costes hasta la capacidad de innovación, pasando por la calidad del capital humano y la infraestructura tecnológica.

Ahora bien, en mi opinión, las irrupciones tecnologías han redefinido los términos en los que se deberían definir ambas características. La automatización de procesos, el análisis masivo de datos, la conectividad permanente y, como disrupción más reciente, la IA generativa, han ampliado claramente los mecanismos a través de los cuales las empresas pueden mejorar su eficiencia y diferenciarse competitivamente. Existen además textos científicos que apoyan este pensamiento; en su libro “Machine Platform Crowd: Harnessing our digital future” Brynjolfsson y McAfee (2017) describen cómo la IA y las nuevas tecnologías son una nueva forma de capital, y cómo las empresas que quieran mantenerse competitivas en esta era digital deberían cambiar sus cadenas de valor y funcionamiento para integrar en sus procesos estos avances.

El impacto de la IA sobre la productividad no se limita a la sustitución de tareas rutinarias por algoritmos, sino que abarca también la reconfiguración de procesos cognitivos complejos, la generación de nuevas formas de conocimiento y la creación de ventajas competitivas

sostenibles basadas en activos intangibles. Sin embargo, aquí hay evidencias que no comparten puntos de vista iguales en cuanto al nivel de impacto y la eficiencia de incluir procesos de IA en el día a día laboral. Zhai y Liu (2023), basándose en empresas chinas cotizadas, muestran que la innovación en tecnologías de IA puede mejorar significativamente la productividad total de los factores, sobre todo en empresas grandes y en industrias intensivas en mano de obra (banca de inversión podría incluirse en este espectro). Por otro lado, hay quienes defienden que existe una “paradoja de la productividad de la IA” (Brynjolfsson et al., 2018), según la cual, debido a las inversiones masivas en un activo intangible como el que suponen estas tecnologías, muchas empresas tardan o no consiguen mejoras de rendimiento pese a la adopción de tecnologías como la IA en sus procesos. Esta división de opiniones, junto con mi experiencia, me hace pensar que el impacto y adopción de la IA dependen también de atributos empresariales como la cultura, la calidad de los sistemas y la formación de los trabajadores humanos.

2.2. Aplicaciones de la IA en banca de inversión

La banca de inversión engloba un conjunto de actividades especializadas que incluyen el asesoramiento en operaciones de fusiones y adquisiciones (M&A), la emisión y colocación de instrumentos financieros en mercados (IPOs o emisiones de bonos), la gestión de carteras de activos, el análisis de mercados y la prestación de servicios de “capital markets”. Estas actividades comparten una serie de rasgos comunes, entre los que para mí destacan la increíble cantidad de datos e información sensible que se analizan y la intensidad y exigencia laborales en las instituciones. Cada operación requiere procesar grandes volúmenes de datos financieros, regulatorios, sectoriales y macroeconómicos, así como elaborar análisis personalizados para clientes institucionales con exigencias de calidad y rapidez excepcionales.

En este contexto, la IA puede tener una aplicación amplia, directa y eficiente en la mayoría de los procesos. En un estudio bibliométrico reciente sobre la integración de la IA en servicios financieros, Goodell et al. (2025) hablan de cómo esta tecnología puede resultar muy útil para la analítica predictiva, la toma de decisiones en tiempo real, la automatización de procesos y la gestión del riesgo. En línea con este argumento, yo mismo, desde mi experiencia a lo largo de la carrera, he visto cómo, a nivel analítico, los modelos de machine learning se pueden utilizar para predecir movimientos en los mercados, valorar activos, analizar oportunidades de arbitraje y realizar análisis de riesgo de crédito en ciertas empresas. Por otro lado, a nivel operativo, la

IA sirve también para automatizar la lectura de contratos, prospectos y documentos legales, y se puede encargar también de redactar informes y materiales de presentación.

Por último, respecto a la relación con clientes, aunque el ser humano pueda parecer que tiene cierta ventaja sostenible sobre la tecnología (es verdad que en general las relaciones humanas y el trato personal son preferidos a interactuar con tecnología), existen ya sistemas de recomendación y asistentes virtuales que facilitan la personalización de productos y ahorran eslabones de jerarquía y cadena de valor a clientes y proveedores (Rabanni et al., 2023).

Todo esto me hace pensar que los modelos de IA generativa abren la puerta a la automatización de tareas tradicionalmente reservadas a analistas junior, como la elaboración de borradores de pitchbooks, la síntesis de informes sectoriales o la generación inicial de modelos de valoración. Se trata de una disrupción que me afectaría de forma directa (tanto a mí como a muchos de mis compañeros universitarios y futuras generaciones), ya sea como una amenaza sustitutiva en mi puesto de trabajo, o como una herramienta clave para ser más eficiente y productivo. Davenport y Ronanki (2018) ya señalaban que la mayoría de las aplicaciones de IA en empresas no buscan sustituir al trabajador, sino aumentar sus capacidades, lo que sugiere que la clave estará en cómo los profesionales del sector aprendan a trabajar con estas herramientas y no contra ellas.

Estos avances tecnológicos sin duda tienen implicaciones directas sobre la productividad del sector y, potencialmente, sobre su estructura organizativa, lo cual justifica la necesidad de estudiar qué se opina a nivel académico y qué tendencias puede haber en un futuro próximo.

2.3. Retos y resistencias a la adopción

Desde mi punto de vista, el sector debería afrontar esta nueva tecnología como una oportunidad, más que como una amenaza. Creo que, tal y como acabo de explicar, la IA generativa puede ser muy útil en ámbitos clave de trabajo en el sector, mejorando mucho la calidad de vida y trabajo de analistas junior que sufren de excesivas horas e intensidad en el modelo actual.

Sin embargo, a pesar del potencial que presenta, la adopción de la IA en banca de inversión tiene también una serie de obstáculos significativos. El primero es de naturaleza regulatoria: el sector financiero está sometido a una serie de normas que exigen documentación constante, capacidad de seguimiento de procesos, explicación detallada en la toma de decisiones, y procesos burocráticos que chocan con el modelo de “caja negra” de muchas IAs avanzadas, donde el usuario no sabe dónde acaban los datos que comparte ni quién está detrás de la

programación del algoritmo (FSB, 2024). El segundo es cultural: la banca de inversión se caracteriza por una elevada aversión al riesgo reputacional y por una fuerte dependencia del juicio humano experto, lo que ralentiza la cesión de decisiones a sistemas automatizados. Al-Dosari et al. (2024) dicen que, aunque las tecnologías de IA ofrecen grandes oportunidades en términos de eficiencia e innovación, la percepción de riesgo por parte de los directivos sigue siendo una barrera relevante en el sector bancario. El tercer riesgo es organizativo: la integración de soluciones de IA requiere transformaciones profundas en los procesos, los flujos de trabajo y las competencias del personal. Y, por último, también existe un cuarto obstáculo ético: el uso de IA en decisiones financieras plantea cuestiones de sesgo, equidad y responsabilidad, más allá del debate sobre la apuesta por IA en el mundo laboral frente a la fuerza humana en términos de desempleo e impacto económico a las personas.

Todos estos factores explican por qué el ritmo de adopción es desigual incluso dentro del propio sector financiero. Un informe reciente del Digital Banking Report (2024) indica que casi una cuarta parte de los bancos aún no ha iniciado ningún proceso de adopción de IA, frente a apenas un 5% de las fintechs. Así pues, considero que entender el estado actual de la investigación académica sobre estas cuestiones es un primer paso importante para comprender las direcciones que puede tomar la transformación del sector en los próximos años en relación con la IA.

3. Metodología

Una vez definidos el contexto y motivación del trabajo, así como el marco teórico en el que se desarrolla este estudio, entramos en el tercer capítulo, donde describiré la metodología empleada para responder a la pregunta de investigación y construir la base de datos que utilizaré posteriormente para sacar conclusiones. El trabajo sigue la estructura de un estudio bibliométrico, que consiste en llevar a cabo una revisión exhaustiva de la literatura científica sobre un tema específico mediante la aplicación de técnicas cuantitativas de análisis de datos. A continuación, explicaré en 6 apartados la metodología que he seguido para llegar a la base final de datos. El capítulo se estructura en: justificación del enfoque bibliométrico, formulación de preguntas y objetivos específicos, fuente de datos, diseño iterativo de la query final de búsqueda de artículos en Scopus, criterios de inclusión y exclusión de los artículos preliminares con el correspondiente diagrama PRISMA, y descripción de las variables exportadas.

3.1. Justificación del enfoque bibliométrico

El análisis bibliométrico es una técnica de análisis cuantitativo que utiliza los datos de artículos científicos (autor, año, fuente, palabras clave, procedencia geográfica de la publicación, el resumen) para entender la estructura, evolución y posibles tendencias dentro un campo de investigación (que en este caso será la IA dentro del mundo laboral y específicamente en banca de inversión). La diferencia de un estudio bibliométrico respecto a un trabajo de investigación tradicional en el que de forma subjetiva y manual se leen todos los artículos científicos que se consideran relevantes, es que este tipo de análisis se apoya en procedimientos sistemáticos y reproducibles que permiten trabajar con volúmenes de estudios y artículos difícilmente manejables de otro modo.

En concreto para este TFG, pienso que el enfoque bibliométrico es especialmente apropiado por tres razones. En primer lugar, nos permite obtener de forma analítica y estructurada una visión general de un fenómeno (la IA) que se encuentra en plena fase de expansión y que es perfectamente aplicable a la disciplina del mundo laboral y la banca de inversión, siendo temas sobre los cuales puede existir literatura científica que examine su compatibilidad (artículos sobre informática, economía, el sector financiero y la inteligencia artificial son compatibles a priori y pueden complementarse). En segundo lugar, a través del análisis bibliométrico, se identifica de forma objetiva a los autores, instituciones, países y revistas que están liderando la investigación en este ámbito. Por último, mediante técnicas de análisis textual en Python aplicadas a títulos, resúmenes y palabras clave, se detectan las principales opiniones y las tendencias futuras sobre la IA en el mundo laboral y en banca de inversión sin necesidad de tener que leer a fondo cada artículo.

La metodología utilizada sigue las propuestas de Öztürk, Kocaman y Kanbach (2024) sobre el diseño de investigaciones bibliométricas, así como las pautas definidas por Passas (2024) acerca de las principales etapas que componen este tipo de estudios.

3.2. Preguntas y objetivos específicos

Según los autores anteriormente mencionados, a raíz de la pregunta y el objetivo generales que se definen al principio de la investigación, conviene también formular preguntas más específicas para dar respuesta al objetivo principal de forma estructurada y adecuada. En línea con este razonamiento, he propuesto seis preguntas específicas, cada una de ellas asociada a un objetivo específico. La formulación de dichas preguntas y sus correspondientes objetivos se basa en: cubrir las principales dimensiones del análisis bibliométrico (evolución temporal,

autores, contenido), mantener la coherencia con el objetivo general, y asegurar la viabilidad técnica del análisis con los datos disponibles en Scopus. La siguiente tabla resume el conjunto:

Pregunta de investigación	Objetivo asociado
PG. ¿Cómo ha evolucionado la investigación científica internacional sobre el impacto de la IA en la productividad y competitividad en el mundo laboral, sobre todo, en banca de inversión, y qué tendencias presenta para el futuro?	OG. Analizar la evolución, estructura y tendencias de la literatura científica internacional sobre el impacto de la IA en la productividad y competitividad en el mundo laboral con foco en banca de inversión.
PE1. ¿Cuál es la evolución del volumen de publicaciones científicas sobre IA aplicada a banca de inversión en los últimos años?	OE1. Entender la evolución temporal de la investigación científica en este ámbito y su grado de consolidación (útil para ver si es algo reciente y emergente o si lleva muchos años siendo estudiado).
PE2. ¿Qué autores, instituciones o países lideran el estudio científico en este ámbito?	OE2. Identificar los principales actores (autores, organizaciones, países) y geografías que tienen el mayor conocimiento de nuestra materia de estudio.
PE3. ¿Qué redes de colaboración y co-autoría existen entre investigadores e instituciones?	OE3. Entender si hay redes de colaboración a nivel internacional al investigar el tema.
PE4. ¿Cuáles son los temas principales tratados en la literatura científica sobre IA y banca de inversión?	OE4. Identificar los temas más relevantes mediante análisis de palabras clave y títulos o resúmenes.
PE5. ¿Qué temas emergentes se observan en los últimos años en este campo?	OE5. Detectar tendencias temáticas emergentes vinculadas a la influencia de la IA en el mundo laboral y la banca de inversión.
PE6. ¿Qué revistas y fuentes aportan la mayoría del estudio en este ámbito?	OE6. Identificar las principales fuentes de publicación y evaluar la dispersión que existe entre autores

Las preguntas PE1, PE2, PE3 y PE6 se centran en las dinámicas de la investigación (quién investiga, dónde, cuándo y a través de qué canales), mientras que las preguntas PE4 y PE5 son preguntas relacionadas con el contenido del tema principal (qué se investiga y hacia dónde está evolucionando ese objeto de estudio, en este caso la IA dentro del mundo laboral y la banca de inversión).

3.3. Fuente de datos: Scopus

La base de datos de artículos científicos que utilizo en mi análisis se ha construido en base a datos obtenidos de la web Scopus, accesible a través de la biblioteca digital de la Universidad Pontificia Comillas ICADE / ICAI. Scopus es, junto con Web of Science, una de las dos principales bases de datos bibliográficas multidisciplinares utilizadas en estudios bibliométricos. El haber elegido Scopus en vez de Web of Science se justifica principalmente por dos motivos: en primer lugar, ofrece una mayor cantidad de información y, en segundo lugar, cada base de datos indexa artículos diferentes, por lo que optar por una sola simplifica el procesamiento y la edición de los datos finales que se obtienen, ya que no se mezclan metadatos de distintas bases.

3.4. Diseño de la query de búsqueda: proceso iterativo de prueba y error

A la hora de encontrar artículos en Scopus, el proceso se hace mediante una query de palabras y términos claves que se construye con términos booleanos ON y AND para tratar de llegar a un número de documentos suficientemente grande, pero a la vez manejable.

De esta forma, el diseño de la query es uno de los pasos más importantes en este proceso bibliométrico, ya que determina tanto la cobertura (capacidad de recuperar todos los artículos relevantes) y la precisión (capacidad de excluir los irrelevantes) del corpus. Una query demasiado amplia introduce ruido y dificulta el filtrado posterior; una query demasiado estrecha deja fuera literatura pertinente y compromete la representatividad del análisis.

Para dar con la query ideal y asegurar una cobertura y precisión correctas (que estimamos en unos 500 documentos para construir la base de datos inicial), he seguido un proceso iterativo de refinamiento de palabras clave. Empezando con una query más amplia que incluía prácticamente todos los términos relacionados con la IA, empresas y productividad, que son de forma general las tres palabras claves del trabajo, he editado progresivamente los términos para centrar la query sobre el foco específico del trabajo, banca de inversión, evaluando en cada paso el volumen de artículo que se obtenía en Scopus con cada búsqueda. Este procedimiento

permite explicar de forma transparente las decisiones que he tomado y justificar la query finalmente elegida. Todas las búsquedas se realizaron sobre los campos Article title, Abstract y Keywords.

La siguiente tabla resume las 6 queries probadas y descartadas hasta llegar a la query definitiva (que expongo a continuación de la explicación del proceso de descarte de estas). En la tabla incluyo detalle de su contenido específico y el número de documentos devuelto por Scopus en cada caso:

N°	Query redactada en el buscador	N° docs.
Q1	("artificial intelligence" OR "machine learning" OR "deep learning") AND ("business" OR "firm" OR "enterprise" OR "company") AND ("productivity" OR "competitiveness" OR "performance")	2.300
Q2	("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*") AND ("finance" OR "financial services" OR "banking" OR "financial institution*") AND ("productivity" OR "competitiveness" OR "efficiency" OR "performance")	1.376
Q3	("artificial intelligence" OR "machine learning" OR "deep learning") AND ("bank*" OR "banking sector" OR "financial institution*") AND ("productivity" OR "competitiveness" OR "efficiency" OR "automation" OR "performance")	771
Q4	("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR "natural language processing") AND ("investment bank*" OR "asset management" OR "portfolio management" OR "capital market*" OR "trading") AND ("productivity" OR "competitiveness" OR "efficiency" OR "performance" OR "automation" OR "decision-making")	2.257
Q5	("artificial intelligence" OR "machine learning" OR "deep learning") AND ("investment bank*" OR "asset management" OR "portfolio management") AND ("productivity" OR "competitiveness" OR "operational efficiency" or "process automation")	18

Q6	(“artificial intelligence” OR “machine learning” OR “deep learning” OR “neural network*”) AND (“investment bank*” OR “asset management” OR “portfolio management” OR “corporate finance” OR “mergers and acquisitions”) AND (“productivity” OR “competitiveness” OR “efficiency” OR “performance” OR “automation” OR “decision-making”)	729
----	---	-----

El proceso para llegar a la query 7 fue el siguiente:

- La Query 1, muy general, devolvió 2.300 documentos sobre IA, empresa y productividad en sentido amplio, sin foco sectorial. Una muestra demasiado amplia y poco enfocada en nuestro objetivo general
- Para la Query 2 introduje más términos de servicios financieros y banca, reduciendo la muestra a 1.376 documentos. Quedó descartada por ser tan amplia, seguramente incluyendo artículos de todos los subsectores que hay dentro del mundo financiero.
- De esta forma, en la Query 3, me centré ya en banca, consiguiendo afinar la búsqueda a 771 documentos. No obstante, la query no distinguía entre el tipo de banca al que nos referíamos (inversión, minorista, etc), por lo que al llegar al filtrado manual en la fase prisma, muchos de los documentos quedaban descartados por no referirse a banca de inversión.
- Para solucionarlo, en la Query 4 incorporé términos específicos de banca de inversión (investment banking, asset management, portfolio management, trading, capital markets), pero de forma inesperada el número de resultados aumentó hasta 2.257: la razón debió ser que los términos trading y capital markets abarcan gran parte de los estudios dentro del sector financiero sobre técnicas de trading algorítmico y modelos cuantitativos. Pese a que la relación de la IA con estas actividades puede ser significativa, son operaciones que no suele incluirse en banca de inversión en sentido estricto (como sí lo son el M&A, los bonos, las acciones o el wealth management).
- En la Query 5 eliminé esos términos, y además ajusté también los términos relacionados con competitividad y eficiencia, para tratar de rebajar aún más el número de artículos. El resultado fue de solo 18 documentos, una muestra insuficiente para nuestro análisis.

- En base a estos resultados construí la Query 6. Eliminando los términos más ruidosos (trading, capital markets) e incluyendo otros más típicos en banca de inversión (corporate finance, mergers and acquisitions), y aumentando los relacionados a la productividad y competitividad, obtuve un total de 729 documentos, un volumen razonable, y a priori enfocado en el tema central, para pasar a los filtros de cribado finales.

No obstante, al pasar a la fase de descarga para hacer una revisión manual preliminar de la base de datos obtenida, se detectó un problema.

El primer paso antes de descargar los documentos obtenidos es fijar los filtros de tipo de documento e idioma en Scopus (en el apartado 3.5 de este trabajo entro en más detalle sobre este proceso). Para la query 6, al aplicar estos filtros, me quedaba con un total de 269 documentos, que son los que después descargué en un csv y filtré de forma manual revisando las palabras clave y título de cada uno para asegurarme de que atendían al foco del trabajo.

Sorprendentemente, tan solo 58 de esos 269 documentos trataban de temas relacionados con IA, productividad y competitividad, y banca de inversión en el sentido financiero. El problema fue que el término “asset management” resultó ser ambiguo en el contexto de Scopus.

El asset managemet que sería relevante para este trabajo es el relacionado a la gestión de activos financieros (carteras de inversión, fondos, acciones). Sin embargo, en mucho de los artículos de Scopus, aparecía como palabra clave de artículos relacionados con la ingeniería y gestión de infraestructuras, donde “asset management” se refiere a la gestión de activos físicos (carretas, puentes, tuberías) durante su ciclo de vida. Lógicamente estos artículos no se relacionan en absoluto con nuestra pregunta central, por lo que la base de datos construida con la query 6 no resultó ser útil.

A raíz de esta reflexión, desarrollé la query 7, que fue la definitiva. En esta última iteración, sustituí el término asset management por otros más específicos a la parte financiera de esa nomenclatura. Además, añadí más palabras clave al bloque de IA basándome en palabras clave que se repetían en los artículos que sí eran de interés con la query 6. Por último, enriquecí también el bloque de productividad y competitividad en el mundo laboral con dos términos más. La query final quedó así:

- *Q7: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR "natural language processing" OR "reinforcement learning")*

AND ("investment bank" OR "investment banking" OR "investment management" OR "financial asset management" OR "wealth management" OR "portfolio management" OR "portfolio optimization" OR "corporate finance" OR "mergers and acquisitions" OR "M&A" OR "stock prediction" OR "stock selection" OR "equity research" OR "mutual fund*" OR "hedge fund*" OR "robo-advisor*")*

AND ("productivity" OR "competitiveness" OR "efficiency" OR "performance" OR "automation" OR "decision-making" OR "innovation" OR "competitive advantage" OR "investment decision" OR "financial decision*")*

Con esta query, se obtuvieron 969 documentos, que, al aplicar el filtro de tipo de documento, se quedó en 338 artículos, siendo esto un volumen razonable y, ahora sí, libre de ruido y sesgos para la siguiente fase de filtración, detallada en el apartado 3.5.

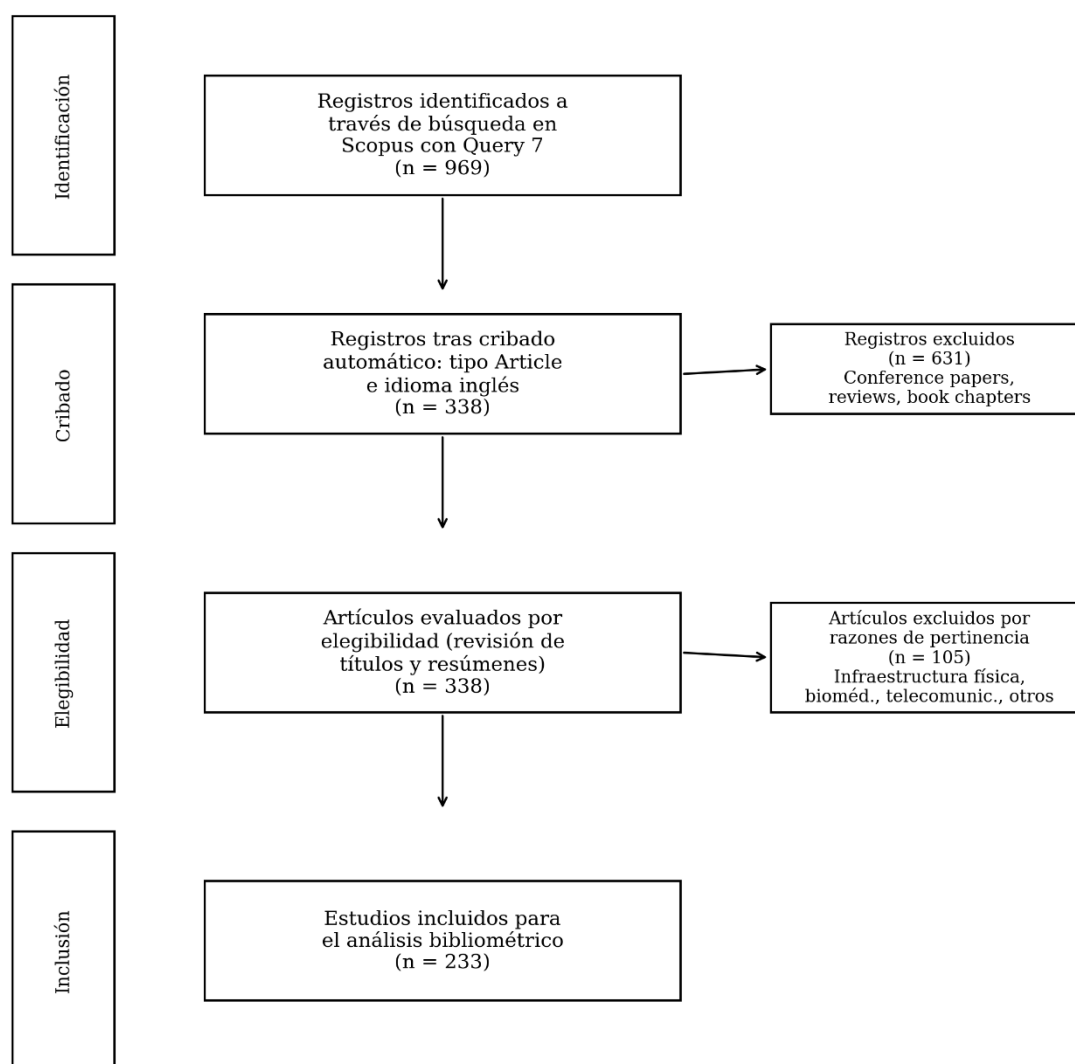
3.5. Proceso de filtrado y diagrama PRISMA

Una vez fijada la query final, se aplicaron una serie de filtros para depurar el corpus que Scopus daba de 969 documentos y obtener una muestra final adecuada al objeto de estudio. Los criterios empleados son los siguientes.

- Tipo de documento: se incluyen únicamente artículos científicos (“Article” dentro del filtro de Scopus), excluyendo conference papers, reviews, o book chapters, Esta restricción garantiza que todos los documentos han pasado por procesos de revisión en revistas indexadas, asegurando cierta calidad académica en nuestros datos.
- Marco temporal: se incluyen todos los documentos publicados hasta 2026. La elección de incluir un eje temporal tan amplio podría ser cuestionada, ya que la IA es un fenómeno de la última década. No obstante, parte del trabajo es estudiar cuándo se ha empezado a tratar este tema académicamente, para ver si es algo reciente, con historia, y cómo de desarrollado está el estudio acerca del tema central.
- Idioma: solo se utilizan publicaciones en inglés, ya que el inglés es considerado el idioma principal a nivel internacional y para la fase de código permite aplicar técnicas estandarizadas de procesamiento de lenguaje natural sobre los artículos (además de que en Scopus, la mayoría de los artículos publicados están en inglés).
- Revisión manual para asegurar coherencia con el tema: tras los filtros automáticos, realicé una revisión manual de títulos y, en caso de duda, resúmenes, para excluir artículos que, pese a estar incluido por la query, no hablan exactamente de la relación

entre IA y banca de inversión (por ejemplo, artículos puramente técnicos sobre arquitecturas de redes neuronales que mencionan finanzas como ejemplo ilustrativo).

El proceso completo de filtrado se puede observar en el siguiente esquema PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)



Se trata de un diagrama estándar en estudios bibliométricos para detallar el proceso de selección de datos. La tabla siguiente resume el flujo de documentos a través de cada fase.

En resumen, los filtros automáticos redujeron los 969 documentos hasta unos 388 artículos. Revisando estos de forma manual por título y resumen, 105 fueron descartados, conservando finalmente un total de 233 artículos para incluir en nuestra base de datos final. Si comparamos la tasa de inclusión en este proceso, que fue del 69% (233 / 388), con la que obtuve con la

query 6, que fue cercana al 20% (58 / 269), se confirma empíricamente la mejora en los resultados del proceso iterativo de elección de las queries y la elección de descartar Q6.

Considero que la base final de 233 artículos es suficientemente grande y manejable para la siguiente fase del proceso, que consistirá en el análisis de los datos mediante técnicas de programación en Python. No obstante, antes de saltar a la fase de análisis y resultados en el capítulo, conviene revisar qué variables he incluido en la base de datos para cada archivo, para así entender del todo los datos iniciales.

3.6. Variables exportadas

Para cada artículo del corpus final, se exportaron los metadatos (variables) disponibles en Scopus en formato CSV, seleccionando los campos relevantes para los objetivos del análisis. La tabla siguiente recoge el conjunto de variables exportadas, agrupadas según los bloques a los que da opción Scopus, dando un breve detalle de por qué pienso que pueden ser útiles:

Bloque Scopus	Variables exportadas	Utilidad
Citation information	Author(s), Document title, Year, EID, Source title, Volume/issues/pages, Citation count, Source & document type, DOI, Open access	Análisis descriptivo, autores y revistas que más salen, evolución temporal
Bibliographical information	Affiliations, Abbreviated source title	Análisis geográfico, redes de colaboración
Abstract & keywords	Abstract, Author keywords, Indexed keywords	Análisis de palabras clave, topic modelling (LDA / BERTopic), filtro manual
Funding details	Funding text	Identificación de agencias y polos de financiación

3.7. Herramientas de análisis

El análisis del corpus se realizará en Python, utilizando la distribución Anaconda y el entorno de desarrollo Spyder. La elección de Python responde a su versatilidad para combinar análisis cuantitativo descriptivo, manipulación de datos en formato tabla y csv, análisis de redes y procesamiento de lenguaje natural. Las técnicas concretas de análisis y los resultados obtenidos se presentan a continuación en el capítulo 4.

4. Análisis de datos y resultados

En esta sección se presenta el análisis con lenguaje de programación Python que llevé a cabo sobre la base final de 233 documentos procedentes de Scopus (el corpus final que construí a raíz de la query 7).

La estructura del análisis se basa en ir dando respuesta a las preguntas específicas de investigación que se detallan en el capítulo 3.2., con el objetivo de, una vez contestadas estas, pasar a dar respuesta a nuestra pregunta general y sacar conclusiones en la última sección del TFG.

Respecto al código empleado, como la principal idea es obtener representaciones gráficas de los datos que nos permitan sacar conclusiones de forma clara y visual, las principales librerías que he utilizado son pandas (con el fin de manejar datos que se encuentran en una tabla), matplotlib (para crear gráficos) y wordcloud (que sirve para generar nubes de palabras, algo que puede ser clave a la hora de analizar documentos con lenguaje natural). También se incluye una sección de topic modelling, en la que analicé más en detalle los abstracts de los artículos para un análisis más riguroso de posibles tendencias y temas principales de estudio. El código completo, correspondiente a los resultados que muestro a continuación, se puede consultar en el anexo 1 del trabajo.

4.1. Evolución temporal del estudio sobre la IA y banca de inversión

La primera pregunta específica que definí (PE1 de la sección 3.2.) tenía como objetivo explicar si el fenómeno que estamos estudiando lleva siendo objetivo de investigación a nivel científico mucho tiempo o si es algo reciente. Una forma sencilla y útil de dar respuesta a esta cuestión es agrupar los documentos de nuestra base de datos según su fecha de publicación (año en este caso).

A continuación, en la figura 4.1, se observa una dinámica clara en cuanto a la distribución temporal del estudio sobre la IA en el mundo laboral y en banca de inversión. Aunque vemos cómo algunos artículos son de finales del siglo XX, la mayor parte de la investigación se ha dado en la última década.

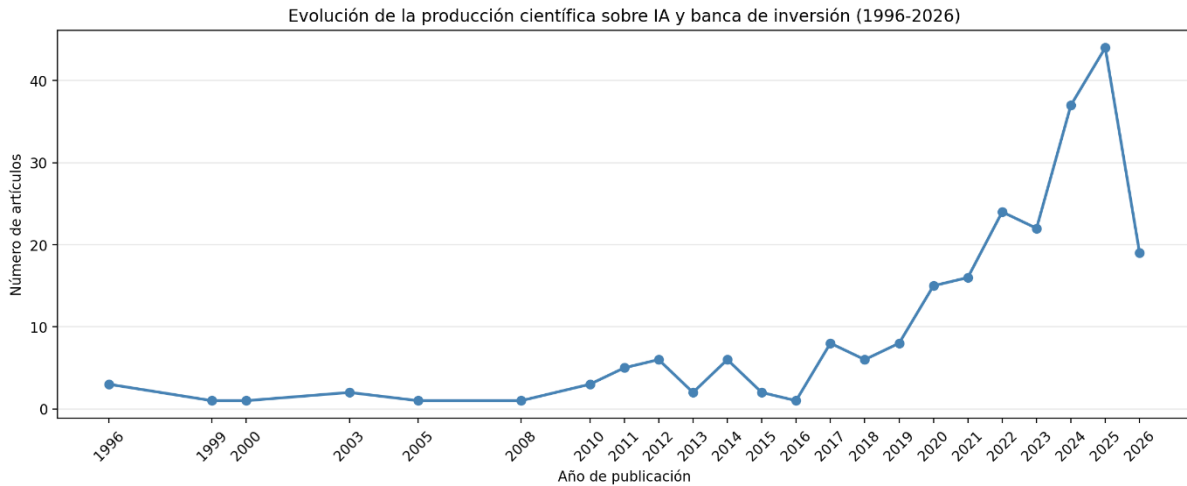


Figura 4.1. Evolución de la producción científica sobre IA y banca de inversión (1996-2026)

Hasta 2017 nunca se superaron los 6 artículos publicados por año. En 2019 es cuando empieza un crecimiento importante de estudio sobre el tema según refleja el gráfico. El máximo número de publicaciones se dio en 2025, con 44 artículos (prácticamente el doble que solo dos años antes).

Esta evolución, que claramente sugiere que el estudio científico en torno a la IA y su aplicación al mundo laboral está en plena fase emergente y de consolidación, tiene sentido, ya que el auge de la IA generativa desde el lanzamiento de ChatGPT a finales de 2022 parece ser justo el momento previo a que se doble la cantidad de estudios publicados acerca de IA y su impacto en el mundo laboral (2024 y 2025).

De esta forma, y en respuesta a la primera pregunta específica del trabajo, considero que el estudio en este ámbito es muy reciente, y todavía debe considerarse como estudio científico en fase de consolidación.

4.2. Autores y países más representativos dentro del ámbito (PE2)

El segundo objetivo específico trata de entender quiénes están liderando la investigación dentro de este campo, en específico, qué autores, instituciones y países.

Para identificar los autores, es clave entender que Scopus agrupa a todos los de un mismo artículo en una única celda, separados por punto y coma (es importante no caer en el fallo de que a un artículo le corresponde un único autor). Por ello, el primer paso fue separarlos antes de poder contarlos individualmente, que lo hice con un bucle que recorre cada artículo, separa la cadena de autores con `split(“;”)` y crea una lista separada con todos los autores. Una vez tengo la lista, conté cuántas veces aparecía cada uno con la función `value_counts()`.

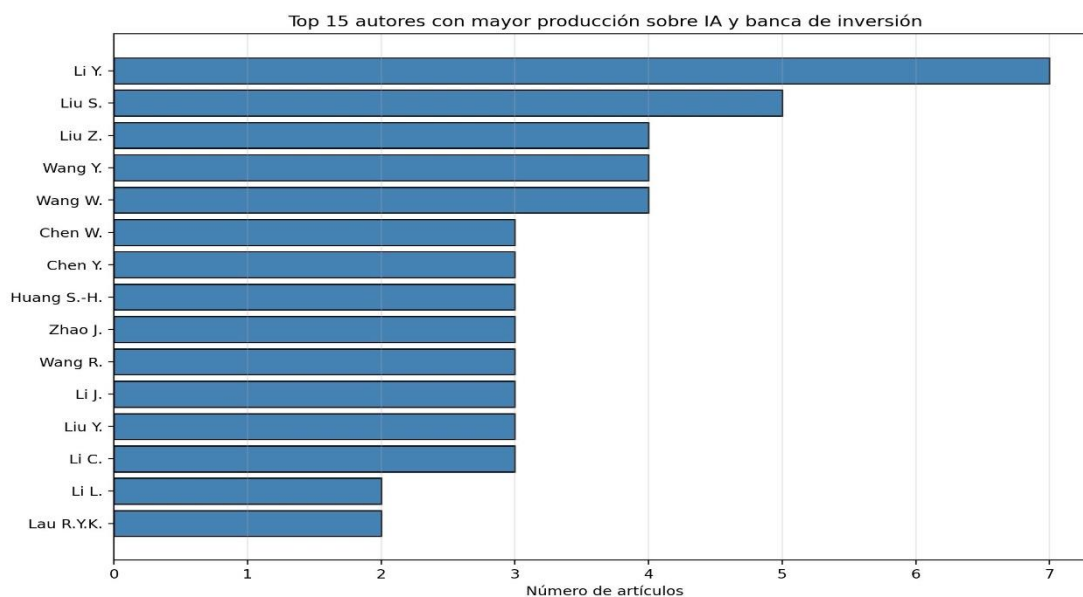


Figura 4.2. Top 15 autores con mayor producción sobre IA y banca de inversión

El autor que más se repite es Li Y., con 7 artículos, seguido por Liu S. (5) y un grupo de autores con 3-4 artículos cada uno. La primera observación es que el estudio no está dominado por una figura individual que destaque de forma considerable por encima del resto: el top 1 representa poco más del 3% del total de artículos publicados e incluidos en nuestra base final. Esto refuerza la idea de que el área es emergente y abierta, sin claros referentes consolidados.

Además, también destaca la gran mayoría de apellidos de origen asiático que vemos en la figura 4.2. obtenida. Esto me hace pensar que la mayoría del estudio acerca del tema se lleva a cabo en geografía asiáticas, algo que investigaré más a fondo en el capítulo 5 de conclusiones para determinar si efectivamente este fenómeno se está cumpliendo y qué razones hay detrás del predominio asiático.

Con respecto a los principales países donde se está llevando a cabo el estudio, viendo los nombres de los autores principales uno puede hacerse a la idea de qué país está liderando la investigación, pero conviene comprobar si efectivamente Asia domina y qué otras regiones son relevantes en este ámbito de estudio.

En este caso, Scopus no da una columna con país específico para cada artículo, sino que los países aparecen dentro del campo de Affiliations, al final de cada bloque y separados por comas. Para extraer el país de cada artículo, cree una lista con los nombres de los principales países en inglés (tal y como aparecen en Scopus), y, para cada artículo, comprobé qué países de esa lista aparecían en su cadena de afiliaciones. Si un país aparecía varias veces en el mismo artículo (porque hay varios autores de ese país), se cuenta solo una vez. Considero que así se evita inflar la representación de países donde se puede trabajar en equipo, centrándome en el país del artículo, más allá del número de autores en ese mismo país de ese artículo.

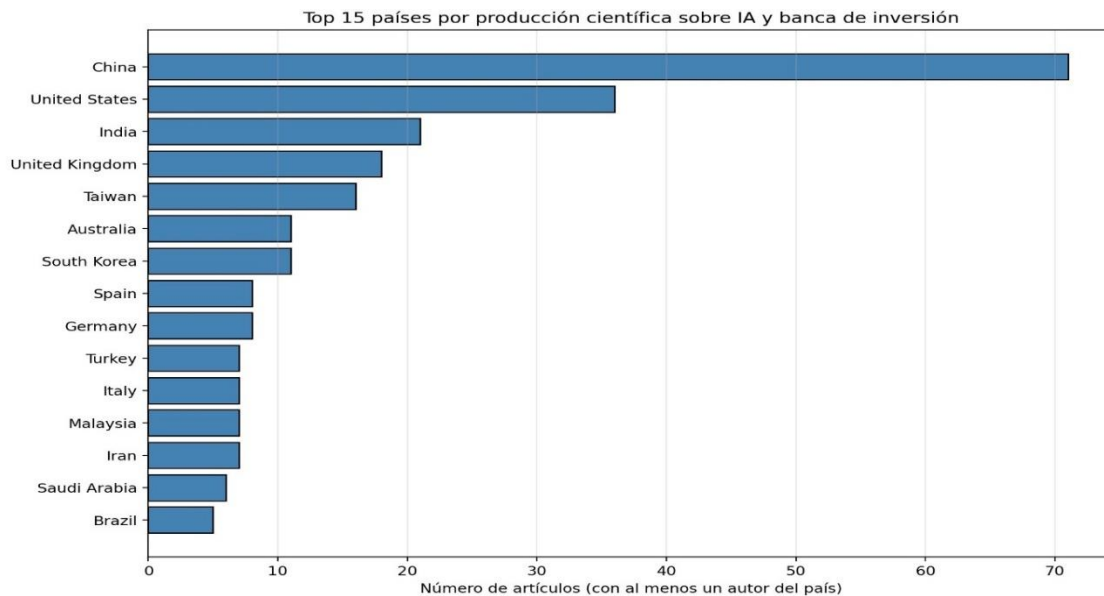


Figura 4.3. Top 15 países por número de artículos publicados

Los resultados confirman la hipótesis del apartado anterior: China lidera la publicación de artículos de forma destacada, con 71 artículos en los que participa al menos un autor con afiliación china (en torno al 30% del total de nuestra base de datos). En segundo lugar, aparece Estados Unidos con 36 artículos (la mitad que China), seguido de India con 21 y Reino Unido con 18. Taiwán, con 16, es otro país llamativo entre los más destacados.

España aparece en octava posición con 8 artículos. Podría ser interesante en futuras investigaciones explorar más a fondo quiénes son esos investigadores españoles y si están vinculados a alguna institución concreta y cuál es su opinión sobre nuestro tema central aplicado a España.

4.3 Temas principales: análisis de palabras clave (PE4)

Una vez tenemos ya una idea general de quién está liderando la investigación tanto a título personal como a nivel geográfico, pasamos a la cuarta pregunta específica que definí en la sección 3 del trabajo: descubrir cuáles son los temas principales dentro de este campo de estudio.

Para ello, analicé la columna de “Author Keywords”, donde cada autor incluye las palabras que considera clave en su artículo. Para el conteo, el procedimiento fue similar al que utilicé en el caso de los autores; separar las palabras clave de cada artículo (vienen separadas por punto y coma), convertirlas todas a minúsculas para evitar duplicaciones (“Machine Learning” y “machine learning” se cuentan como el mismo término) y contar la frecuencia de cada palabra.

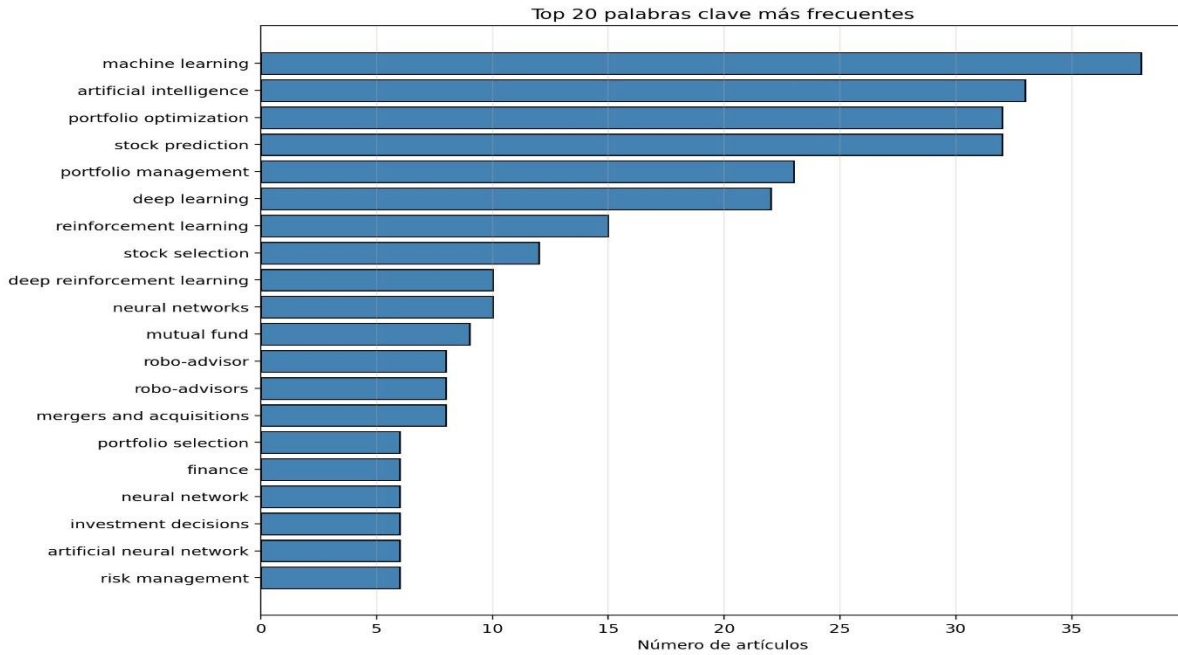


Figura 4.4. Top 20 palabras clave más frecuentes

A la hora de analizar estos resultados, creo que es útil dividirlos en dos partes: por un lado, los términos que se refieren a inteligencia artificial y por otro, los relacionados con banca de inversión, y a raíz de lo observado en cada uno, comparar y sacar conclusiones.

Respecto a la parte de IA, muchos de los términos encabezan el ranking que obtuve en este primer análisis: machine learning (38 artículos), artificial intelligence (33), deep learning (22), reinforcement learning (15), deep reinforcement learning (10) y neural networks (10). Se trata de términos de modelos complejos de redes neuronales profundas y de aprendizaje por refuerzo, en línea con los algoritmos que soportan los modelos emergentes de IA.

Por su parte, respecto al apartado financiero, los términos más repetidos son portfolio optimization (32), stock prediction (32), portfolio management (23) y stock selection (12). Es decir, la mayor parte de la investigación se centra en problemas de gestión de carteras y predicción de precios de acciones o bonos. Otros términos también aparecen con frecuencia notable pero menor: mutual fund (9), mergers and acquisitions (8), investment decisions (6) y risk management (6).

Esto me hace pensar que, aunque mi pregunta principal se centra en el impacto de la IA en las actividades de banca de inversión más tradicionales de M&A, valoración de empresas, generación de pitchbooks o análisis sectoriales, los datos muestran que la literatura académica está mucho más volcada en aplicaciones cuantitativas de gestión de carteras y predicción bursátil (algo que se incluye en banca de inversión, pero donde no he tenido la experiencia personal que describía en el marco teórico y la introducción de este trabajo). Así pues, se

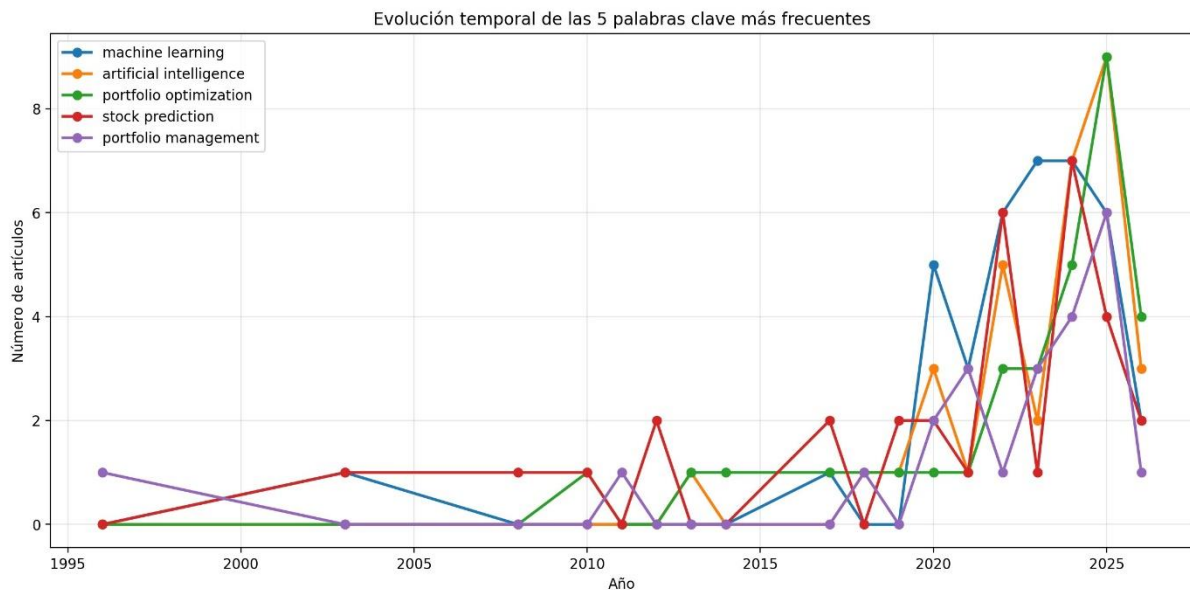


Figura 4.6 Evolución temporal de las 5 palabras clave más frecuentes

Como era de esperar, el gráfico revela patrones similares a la evolución general de la investigación que se vio en el apartado 4.1. Hasta aproximadamente 2018, las cinco palabras clave aparecen de forma casi esporádica (1-2 artículos por año como mucho). A partir de 2019-2020 todas empiezan a crecer, y a partir de 2022 se observa una explosión clara, con picos en 2024 y 2025.

Pese a que este análisis puede resultar útil, y refuerza el hecho de que estamos ante un fenómeno en plena consolidación y con amplio recorrido de evolución por delante, es cierto que no se pueden sacar tendencias futuras sobre la relación de IA y banca de inversión muy concluyentes con este gráfico.

Por ello, decidí llevar a cabo un ejercicio de topic modelling (explicado en el siguiente apartado), donde entro de forma más profunda en los abstracts completos de los artículos que tengo para tratar de analizar estos temas mejor y ver si a raíz de ello hay alguna tendencia clara más allá de las que ya hemos descubierto.

4.5 Topic modelling (PE4 + PE5)

El análisis sobre palabras clave en los apartados 4.5 y 4.6 ofrece una primera aproximación a qué temas son los más relevantes y en qué dirección se orienta la investigación sobre IA en banca de inversión a día de hoy. No obstante, estas palabras pueden ser términos generales, que los autores eligen para describir sus artículos y no reflejar con precisión de qué trata en realidad el artículo si uno lo lee en profundidad.

Para completar ese análisis, apliqué una técnica de modelado temático llamada BERTopic sobre los 233 abstracts (resúmenes) de la base de datos final. BERTopic es un algoritmo que lee el texto del resumen, lo convierte en una lista larga de números con cientos de posiciones (por ejemplo, algo como 0.23, -0.87, 0.45, 1.12...), y luego agrupa los artículos que hablan de temas similares (es decir, agrupa las listas que tengan números parecidos bajo un único topic, como si al estar en un plano estuvieran cerca sus coordenadas). A diferencia del análisis que había hecho hasta ahora, BERTopic no depende de lo que los autores definen como términos claves en sus estudios. Para cada tema identificado (“topic x”), el modelo devuelve las 10 palabras más representativas, lo que permite interpretar de qué trata cada grupo de palabras. El modelo está configurado con un tamaño mínimo de 10 artículos por tema (para evitar temas residuales con pocos artículos) y en teoría buscaba mínimo devolver hasta 8 temas. Para ser correcto, tuve que eliminar las stopwords en inglés y detectar bigramas. Una vez ejecutado el código, BERTopic identificó tan solo dos temas principales, expuestos en el gráfico a continuación, junto con las 3 principales palabras dentro de las 10 que definen estos topics:

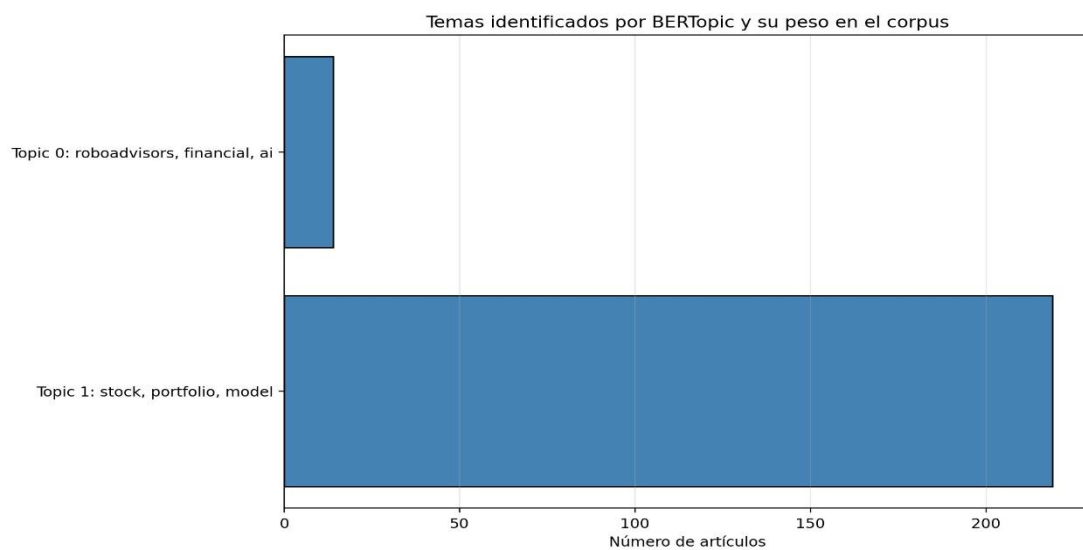


Figura 4.7 Topics identificados por BERTopic en el corpus final

Pese a que en el código traté de forzar al algoritmo a encontrar 8 topics, el resultado final fue tan solo dos, indicando cierta homogeneidad de los artículos en cuanto a contenido temático, algo que analizaré con más detalle más adelante para determinar si es un resultado positivo, negativo o qué implicaciones tiene.

El tema dominante (**Topic 1**, 219 artículos, 94% del corpus) se define por las palabras: **stock, portfolio, model, learning, performance, market, data, financial, optimization, models**. Considero que se trata del núcleo central de la investigación: artículos que tratan sobre técnicas

de machine learning y deep learning aplicadas a la predicción de rendimientos de mercados y a la optimización de carteras de inversión. Volvemos a ver cómo se trata de temas cuantitativos, que se centran en el desarrollo y comparación de algoritmos para mejorar estrategias de inversión.

El otro topic devuelto por el modelo (**Topic 0**, 14 artículos, el 6% del corpus) incluye como palabras clave: **roboadvisors, financial, ai, perceived, study, service, advice, intelligence, resistance, investment**. Se trata de un tema cualitativamente distinto: estos artículos se centran en los robo-advisors y en la percepción, adopción y resistencia de los usuarios frente a los servicios financieros basados en IA. Es curioso que ahora vemos palabras como perceived, resistance y advice, que indican la existencia de artículos que estudian las barreras de adopción que puede haber de la IA dentro de otros servicios financieros y la opinión sobre su evolución, en lugar de estudiar el desarrollo de modelos técnicos para el trading.

El hecho de que el algoritmo solo haya identificado dos temas no implica que algo haya fallado. Al revés, considero que se trata de un hallazgo importante y en línea con otros que venía comentando. Obteniendo solo dos topics, y siendo uno de ellos tan pesado frente al otro, podemos concluir que el corpus es bastante homogéneo temáticamente y también algo acerca de la investigación en este ámbito: la gran mayoría de la investigación converge en un mismo paradigma (aplicar modelos de IA a problemas cuantitativos de inversión), y solo un nicho reducido aborda la dimensión humana y de adopción de estas herramientas.

Es útil también ver una evolución temporal de estos tópicos, para así descubrir hacia donde está yendo la investigación; si va a seguir siendo sobre modelos de predicción, o si por el contrario cada vez está ganando más peso el estudio sobre el debate de la adopción de la IA en el mundo financiero.

Para ello, construí un gráfico de líneas de evolución temporal en el que se puede observar la evolución de ambos tópicos:

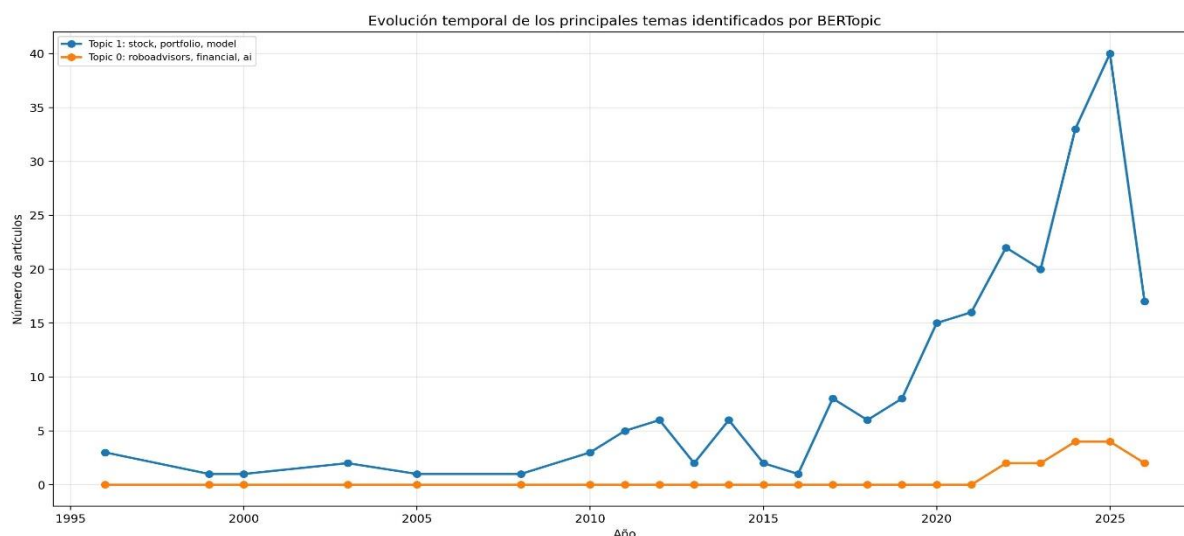


Figura 4.8 Evolución temporal de topics obtenidos con BERTopic

El Topic 1 (stock/portfolio/model) experimentó una fuerte explosión en cuanto a peso en el estudio a partir del año 2020, culminando con un total de 45 artículos en 2025. Se trata de un tema que ha tenido mayor peso en la investigación incluso desde 2010 en adelante, volviendo a reforzar que la IA en finanzas se asocia con modelos de trading predictivo. El Topic 0 (roboadvisors/financial/advice), en cambio, no aparece hasta 2020 y su crecimiento es todavía pequeño, con un máximo de 4 artículos en 2024. Se trata, por tanto, de un tema emergente, pero que si observamos el gráfico, parece estar ganando peso, algo que podría prolongarse en los próximos años a medida que los robo-advisors se generalicen y la investigación empiece a interesarse más por cómo los profesionales y clientes del sector financiero interactúan con estas herramientas de IA (algo que va muy en línea con y sería muy interesante para nuestro objetivo central de investigación).

Como conclusión de este ejercicio de topic modelling, pienso que el resultado obtenido es positivo para nuestra línea de investigación. La banca de inversión es un sector donde las resistencias culturales, regulatorias y organizativas son tan relevantes como las capacidades técnicas de la IA. Sin embargo, este análisis temático muestra que se ha dedicado la inmensa mayoría del estudio a la dimensión técnica (modelos de predicción y optimización) y muy poco a la dimensión humana y de adopción. Esto sugiere un vacío de investigación significativo que futuras contribuciones podrían abordar y una tendencia emergente que ya observamos en los últimos años de artículos empezando a focalizarse en este aspecto más cualitativo de la relación entre IA y banca de inversión en el día a día laboral.

4.6 Revistas con mayor producción (PE6)

Por último, y para dar respuesta a la sexta y última pregunta específica, construí un nuevo gráfico de barras con el que podemos ver de forma sencilla cuáles son las revistas que lideran toda esta investigación hoy en día. Utilizando *value_counts()* sobre la columna de Source title de nuestra base final de datos, me quedé con las 15 revistas más frecuentes entre los artículos publicados.

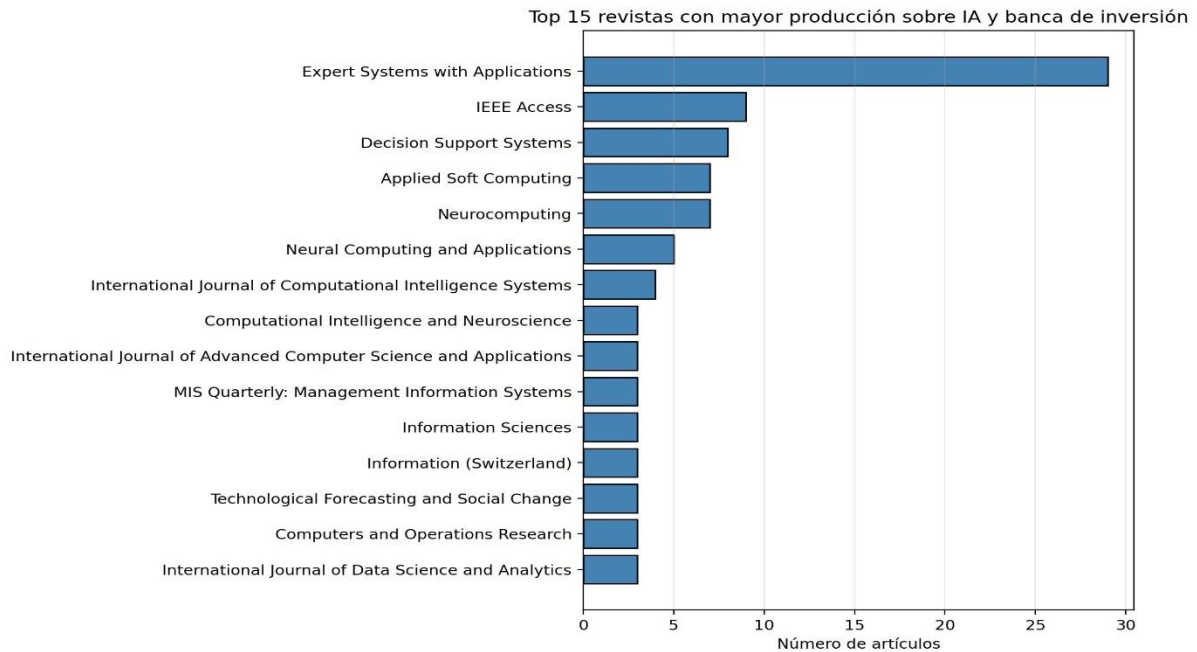


Figura 4.9 Top 15 revistas con mayor número de publicaciones sobre IA y banca de inversión

En este caso, los resultados reflejan algo muy interesante: existe una clara concentración de publicaciones que pertenecen a la misma revista, *Expert Systems with Applications*, que con 29 artículos acumula el 12,4% de toda la producción del corpus. La distancia con la segunda revista (*IEEE Access*, con 9 artículos) y la tercera (*Decision Support Systems*, con 8) es muy notable. A partir del puesto 6, todas las revistas tienen entre 3 y 7 artículos, lo que indica una larga cola de fuentes con una representación minoritaria. Esto indica que pese a ser un ámbito de estudio en fase emergente, según la base de datos que he construido, este campo de investigación parece tener ya cierta estructura editorial.

Pero lo más interesante en este apartado se descubre al ver qué tipo de revistas son estas que hemos obtenido. Casi todas son revistas técnicas de informática, ingeniería o ciencias de la decisión (*Expert Systems with Applications*, *IEEE Access*, *Decision Support Systems*, *Applied Soft Computing*, *Neurocomputing*, *Neural Computing and Applications*). Las revistas específicas de finanzas o gestión empresarial son una minoría dentro del top 15. Esto quiere decir que la investigación sobre IA en banca de inversión se lleva a cabo principalmente desde el lado de la informática y no tanto desde el lado de las finanzas. Es algo importante de cara a

este TFG: implica que probablemente los avances técnicos van por delante de la reflexión académica desde la perspectiva del sector financiero, lo cual encajaría bastante con lo que comenté en el marco teórico sobre la lentitud de adopción de la IA en banca de inversión.

Con todo esto, concluyo este capítulo 4, para a continuación, en la sección 5, discutir las principales conclusiones que saco de este análisis, las limitaciones que me he encontrado en mi ejercicio, y futuras líneas de investigación sobre el tema central de este trabajo.

5. Conclusiones, limitaciones y futuras líneas de investigación

Los resultados obtenidos en el capítulo anterior permiten responder a la pregunta general de investigación y, además, también sirven para profundizar en ciertos descubrimientos para buscar el porqué de estos resultados más allá de lo que salió con el código.

5.1. Discusión de los resultados obtenidos

En este apartado aprovecho para reflexionar en mayor medida sobre algunos de los resultados anteriores y opinar sobre lo que pueden significar para el campo de estudio sobre este tema y para la relación entre IA y banca de inversión.

El primer resultado que me parece relevante comentar es la evolución temporal de la investigación científica en estos ámbitos. Se trata de un tema de estudio joven: hasta 2017 casi no se publicaban artículos sobre IA aplicada a banca de inversión, y la explosión llega a partir de 2019-2020. Esto coincide con dos fenómenos tecnológicos: el boom del deep learning (2017-2018) y el lanzamiento de ChatGPT y la IA generativa (finales de 2022). Creo que esta coincidencia no es algo casual. Los investigadores científicos estudian lo que más impacto tiene en la sociedad, y el interés que ha surgido a raíz de la IA en el mundo laboral y en finanzas es un reflejo directo del impacto mediático y empresarial que estas tecnologías han tenido en los últimos años. Lo que aún no sabemos es si este crecimiento se mantendrá o si se trata de un pico de hype de la IA generativa. Pero el hecho de que justo en el momento en el que surgen estas tecnologías, se dispare la producción científica acerca de ellas y en el sector financiero, refleja el impacto e importancia del objeto de estudio de este TFG.

En segundo lugar, los hallazgos de distribución geográfica también son interesantes y dan que pensar. China es líder con un 30% de aportación, doblando a Estados Unidos y triplicando a India y Reino Unido. En conjunto, el bloque de países asiáticos contribuye mucho más que el bloque occidental de Estados Unidos y Europa. Esto me resulta llamativo y, en cierto modo, paradójico. Los grandes centros de banca de inversión del mundo están en Nueva York y

Londres, no en Pekín o Bombay. ¿Por qué entonces la investigación académica sobre IA en este sector está liderada por Asia? Una posible explicación es que las universidades asiáticas, especialmente las chinas, son muy fuertes en investigación en IA y machine learning, y aplican esas técnicas a problemas financieros. En cambio, las universidades occidentales que investigan sobre banca de inversión probablemente lo hacen desde departamentos de finanzas o gestión, donde las herramientas de IA no están consolidadas, ya que los métodos de investigación instaurados vienen de muchos años atrás. Aunque requiere más estudio a fondo para ver si estas hipótesis explican los resultados obtenidos en este TFG, sí podemos afirmar que parece haber una desconexión entre quienes desarrollan las herramientas (investigadores de IA en Asia) y quienes podrían beneficiarse de ellas (profesionales de banca en Occidente), lo que podría explicar en parte la lentitud de adopción de IA en banca que he observado durante mis prácticas en distintas instituciones.

Por último, el apartado de análisis de palabras clave y de modelado temático refuerzan que a día de hoy, el estudio se concentra de forma abundante en prácticas de portfolio optimization, stock prediction, portfolio management o reinforcement learning. Utilizando BERTopic se ve que el 94% del corpus está centrado en modelos de predicción de mercados y optimización de carteras. Solo un 6% se ocupa de un tema distinto: los robo-advisors y la percepción de los servicios financieros con IA (advice, adoption, etc). Este descubrimiento me parece muy interesante y me da pie a reflexionar y opinar desde mi experiencia personal: durante mis prácticas en banca de inversión, las tareas que más tiempo consumían a los analistas junior no eran de predicción de acciones ni de gestión de carteras, sino de modelos financieros en Excel, preparación de materiales en PowerPoint, análisis de documentos de due diligence y análisis de sectores y riesgos macroeconómicos. Estas tareas, que son precisamente las que más podrían beneficiarse de la IA generativa para ahorrar tiempo y aportar eficiencia y ventajas competitivas en el sector, son prácticamente nulas entre los términos principales de la literatura académica que hemos analizado. Hay, por tanto, un vacío significativo entre lo que se investiga y lo que yo pienso que el sector podría utilizar para ser más eficiente.

Dicho esto, el análisis de la evolución temporal del Topic 0 (robo-advisors y adopción) muestra que este tema no existía antes de 2020 y está creciendo en los últimos años. Esto me da una visión más optimista: parece que la literatura sobre la dimensión humana y de adopción de la IA en servicios financieros está empezando a ganar popularidad y peso dentro del mundo científico. Si esta tendencia se mantiene, es posible que en los próximos años veamos más estudios sobre cómo los profesionales del sector perciben, adoptan y utilizan herramientas de IA en su trabajo diario.

5.2. Limitaciones del trabajo

Pese a todos los aprendizajes, y la utilidad de esta investigación, este trabajo tiene una serie de limitaciones que conviene resaltar, para que la interpretación de los resultados se haga en el contexto adecuado.

En primer lugar, respecto a la fuente de datos, el corpus se ha construido exclusivamente a partir de Scopus. Esto supone una restricción, ya que, aunque Scopus ofrece mayor cobertura que Web of Science en este caso, es posible que artículos relevantes no incluidos en Scopus u otras bases hayan quedado fuera del análisis. Un trabajo en el que se combinan ambas fuentes habría aportado mayor cobertura y riqueza de artículos, aunque también es verdad que podría dar pie a problemas de duplicación o ruido, que podrían empeorar la calidad de los resultados obtenidos.

En segundo lugar, existen también limitaciones respecto al diseño de la query. Aunque llevé a cabo un proceso iterativo riguroso, incluso cambiando la query 6 tras haber decidido que podía ser la definitiva, creo que la naturaleza de los términos implica que siempre va a haber ruidos o ciertos artículos que sí eran relevantes y por no incluir un término clave los hemos pasado por alto.

En tercer lugar, y una de las más relevantes a mi parecer, es el tamaño del corpus. Con 233 artículos, la muestra es válida para el análisis descriptivo y de palabras clave, pero limita un poco la capacidad de técnicas más avanzadas como el topic modelling. Con el algoritmo BERTopic solo pude identificar dos temas, lo que refleja tanto la homogeneidad del corpus como las limitaciones de la técnica con muestras de este tamaño. Un corpus más amplio probablemente habría permitido sacar conclusiones más amplias.

Por último, creo que el análisis bibliométrico ha sido muy útil para entender la evolución del estudio acerca de la IA y banca de inversión, cuáles son las tendencias generales hoy en día y qué términos son clave en este análisis, pero el trabajo se basa en metadatos (título, abstracts, palabras clave, afiliaciones), no en el texto completo de los artículos. Considero que para estudiar mejor las conclusiones que existen hoy en día sobre la posible entrada de la IA en banca de inversión haría falta leer a fondo el texto de los 14 artículos que sacamos en el topic 0 en el ejercicio con BERTopic (aquellos que hablaban más de la relación cualitativa entre IA y banca).

5.3. Futura líneas de investigación

Ahora bien, a partir de ese análisis de limitaciones, y basándome en mis hallazgos con este trabajo, creo que existen varias líneas de investigación futura que pueden ser relevantes e interesantes.

Creo que, tal y como comentaba al final del apartado anterior, una siguiente forma de investigación debería consistir en el análisis más a fondo del contenido como tal de los artículos específicos sobre la adopción de la IA en los bancos de inversión, atendiendo a barreras y promotores de la adopción, para entender si se trata de una opción real para el futuro.

Además, otra forma de complementar este estudio puramente bibliométrico sería con entrevistas a profesionales de banca de inversión, para contrastar si los resultados obtenidos a través de metadatos cuadran con la percepción de aquellos que más experiencia tienen en el sector y viven estos cambios en el día a día.

Por último, creo que sería interesante repetir este mismo análisis dentro de dos o tres años para comprobar si la tendencia emergente del Topic 0 (robo-advisors y adopción) se consolida y la IA generativa se convierte en un tema central de la investigación sobre banca de inversión.

5.4. Reflexión final

Este TFG surgió en gran parte a raíz de mi experiencia personal durante mis prácticas en banca de inversión: creo que la IA podría transformar profundamente el sector para bien, pero que su adopción avanza con demasiada lentitud y precaución.

El análisis bibliométrico realizado confirma que la investigación académica sobre el tema es más bien reciente, está en fase de expansión y se centra sobre todo en aplicaciones cuantitativas de predicción de rendimientos de mercado y carteras de inversión.

Las tareas que yo llevé a cabo y en las que creo que la IA podría mejorar la eficiencia y competitividad dentro del sector de banca de inversión son de un carácter más cualitativo y manual. En este trabajo he aprendido que la posible relación e impacto de la IA sobre estas está siendo estudiada en menor medida. Al principio del trabajo pensaba que esto ni si quiera estaba siendo estudiado, pero con el análisis de Topic Modelling, en concreto con la agrupación del Topic 0 como segundo tema principal, y su evolución que ha ido ganando más peso en los últimos años, pienso que es un tema que se está empezando tomar más en serio y puede llegar a ser muy relevante en un futuro no tan lejano.

La banca de inversión no necesita solo mejores algoritmos de predicción: necesita entender cómo integrar la IA en sus procesos, en su cultura y en la relación con sus clientes. Ese es, en mi opinión, el verdadero reto para los próximos años.

6. Bibliografia

Brynjolfsson, E., & McAfee, A. (2017). *Machine, Platform, Crowd: Harnessing Our Digital Future*. <https://www.futureforwork.com/assets/uploads/2018/10/Rese%c3%b1as-FFWi-N%c3%bamera-003.pdf>

Zhai, S., & Liu, Z. (2023). Artificial intelligence technology innovation and firm productivity. *Finance Research Letters*, 58, 104397
<https://www.sciencedirect.com/science/article/abs/pii/S1544612323008097>

Brynjolfsson, E., Rock, D., & Syverson, C. (2018). *The productivity J-curve: How intangibles complement general purpose technologies* (NBER Working Paper No. 25148). National Bureau of Economic Research
https://www.nber.org/system/files/working_papers/w25148/w25148.pdf

Goodell, J. W., Kumar, S., Li, X., Pattnaik, D., & Sharma, A. (2025). AI integration in financial services: A systematic review of trends and regulatory challenges. *Humanities and Social Sciences Communications*, 12, 450. <https://www.nature.com/articles/s41599-025-04850-8>

Rabbani, M. R., Khan, S., & Thalassinou, E. I. (2023). FinTech, blockchain and Islamic finance: An extensive literature review. *International Journal of Economics and Business Administration*, 8(2), 65–86 <https://ijeba.com/journal/444>

Davenport, T., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116

Financial Stability Board. (2024). *The financial stability implications of artificial intelligence*. <https://www.fsb.org/uploads/P14112024.pdf>

Al-Dosari, K., Fetais, N., & Kucukvar, M. (2024). Adoption of artificial intelligence and machine learning in banking systems: A qualitative survey of board of directors. *Frontiers in Artificial Intelligence*, 7, 1440051 <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1440051/full>

Digital Banking Report. (2024). *The state of AI in banking*.

<https://www.opentext.com/media/report/state-of-ai-in-banking-digital-banking-report-en.pdf>

7. Anexo: código de Python utilizado en la sección 4

```
import pandas as pd
import matplotlib.pyplot as plt

# Cargar el corpus
df = pd.read_excel("corpus_final_completo.xlsx")

# Comprobaciones básicas
print("Número de artículos:", len(df))
print("Número de variables:", len(df.columns))
print("\nPrimeras columnas:", list(df.columns)[:8])
print("\nRango temporal:", df['Year'].min(), "-", df['Year'].max())

# =====
# PARTE 1: Evolución temporal (PE1)
# =====

# Contar publicaciones por año
pubs_por_año = df['Year'].value_counts().sort_index()
print("\nPublicaciones por año:")
print(pubs_por_año)

# Gráfico de líneas
plt.figure(figsize=(12, 5))
plt.plot(pubs_por_año.index, pubs_por_año.values, color='steelblue', marker='o', linewidth=2)
plt.xlabel('Año de publicación')
plt.ylabel('Número de artículos')
plt.title('Evolución de la producción científica sobre IA y banca de inversión (1996-2026)')
plt.xticks(pubs_por_año.index, rotation=45)
plt.grid(axis='y', alpha=0.3)
```

```

plt.tight_layout()
plt.savefig('grafico_evolucion_temporal.png', dpi=200, bbox_inches='tight')
plt.show()

# =====
# PARTE 2: Top autores (PE2)
# =====

# Separar autores: cada artículo tiene varios autores juntos en una celda
todos_autores = []
for autores_str in df['Authors'].dropna():
    autores = [a.strip() for a in autores_str.split(';')]
    todos_autores.extend(autores)

# Contar
top_autores = pd.Series(todos_autores).value_counts().head(15)
print("\nTop 15 autores por número de artículos:")
print(top_autores)

# Gráfico
plt.figure(figsize=(10, 7))
plt.barh(top_autores.index[::-1], top_autores.values[::-1], color='steelblue', edgecolor='black')
plt.xlabel('Número de artículos')
plt.title('Top 15 autores con mayor producción sobre IA y banca de inversión')
plt.grid(axis='x', alpha=0.3)
plt.tight_layout()
plt.savefig('grafico_top_autores.png', dpi=200, bbox_inches='tight')
plt.show()

# =====
# PARTE 3: Top países (PE2)
# =====

```

```

# Lista de países que esperamos encontrar en las afiliaciones
# (esto evita falsos positivos al extraer)
paises_conocidos = [
    'United States', 'China', 'United Kingdom', 'India', 'Germany', 'Spain',
    'Italy', 'France', 'Australia', 'Canada', 'Netherlands', 'Switzerland',
    'Japan', 'South Korea', 'Singapore', 'Hong Kong', 'Taiwan', 'Brazil',
    'Mexico', 'Portugal', 'Belgium', 'Sweden', 'Norway', 'Denmark', 'Finland',
    'Greece', 'Poland', 'Russia', 'Turkey', 'Saudi Arabia', 'United Arab Emirates',
    'Iran', 'Pakistan', 'Bangladesh', 'Vietnam', 'Indonesia', 'Malaysia',
    'Thailand', 'Philippines', 'South Africa', 'Egypt', 'Israel', 'Ireland',
    'Austria', 'Czech Republic', 'Romania', 'Ukraine', 'New Zealand', 'Chile',
    'Argentina', 'Colombia', 'Peru'
]

# Recorremos cada artículo y buscamos qué países aparecen en sus afiliaciones
todos_paises = []
for afiliaciones in df['Affiliations'].dropna():
    paises_articulo = set() # set para evitar contar el mismo país 2 veces en un mismo artículo
    for pais in paises_conocidos:
        if pais in afiliaciones:
            paises_articulo.add(pais)
    todos_paises.extend(list(paises_articulo))

# Contar
top_paises = pd.Series(todos_paises).value_counts().head(15)
print("\nTop 15 países por número de artículos:")
print(top_paises)

# Gráfico
plt.figure(figsize=(10, 7))
plt.barh(top_paises.index[:-1], top_paises.values[:-1], color='steelblue', edgecolor='black')
plt.xlabel('Número de artículos (con al menos un autor del país)')
plt.title('Top 15 países por producción científica sobre IA y banca de inversión')
plt.grid(axis='x', alpha=0.3)

```

```

plt.tight_layout()
plt.savefig('grafico_top_paises.png', dpi=200, bbox_inches='tight')
plt.show()

# =====
# PARTE 4: Análisis de palabras clave (PE4)
# =====

# Separar las palabras clave (vienen separadas por ;)
todas_keywords = []
for kw_str in df['Author Keywords'].dropna():
    keywords = [k.strip().lower() for k in kw_str.split(';')]
    todas_keywords.extend(keywords)

# Contar
top_keywords = pd.Series(todas_keywords).value_counts().head(20)
print("\nTop 20 palabras clave:")
print(top_keywords)

# Gráfico
plt.figure(figsize=(10, 8))
plt.barh(top_keywords.index[::-1], top_keywords.values[::-1], color='steelblue',
edgecolor='black')
plt.xlabel('Número de artículos')
plt.title('Top 20 palabras clave más frecuentes')
plt.grid(axis='x', alpha=0.3)
plt.tight_layout()
plt.savefig('grafico_top_keywords.png', dpi=200, bbox_inches='tight')
plt.show()

# =====
# PARTE 5: Nube de palabras (complemento PE4)
# =====

```

```

from wordcloud import WordCloud

# Convertir el conteo de keywords en un diccionario {palabra: frecuencia}
kw_freq = pd.Series(todas_keywords).value_counts().to_dict()

# Generar la nube
wc = WordCloud(width=1200, height=600, background_color='white',
               colormap='Blues',
               prefer_horizontal=0.9).generate_from_frequencies(kw_freq,
                                                               max_words=80)

plt.figure(figsize=(14, 7))
plt.imshow(wc, interpolation='bilinear')
plt.axis('off')
plt.title('Nube de palabras clave del corpus', fontsize=14)
plt.tight_layout()
plt.savefig('nube_keywords.png', dpi=200, bbox_inches='tight')
plt.show()

# =====
# PARTE 6: Evolución temporal de keywords (PE5)
# =====

# Cogemos las 5 palabras clave más frecuentes
top5_kw = top_keywords.head(5).index.tolist()
print("\nAnalizando evolución de:", top5_kw)

# Para cada artículo, sacamos sus palabras clave y su año
# y contamos cuántas veces aparece cada palabra del top 5 en cada año
evolucion = {kw: {} for kw in top5_kw}

for _, row in df.iterrows():
    if pd.isna(row['Author Keywords']):
        continue
    año = row['Year']

```

```

kws_articulo = [k.strip().lower() for k in row['Author Keywords'].split(';')]
for kw in top5_kw:
    if kw in kws_articulo:
        evolucion[kw][año] = evolucion[kw].get(año, 0) + 1

# Convertir a DataFrame para graficar
evol_df = pd.DataFrame(evolucion).fillna(0).sort_index()
print("\nEvolución temporal:")
print(evolver_df)

# Gráfico de líneas
plt.figure(figsize=(12, 6))
for kw in top5_kw:
    plt.plot(evolver_df.index, evol_df[kw], marker='o', label=kw, linewidth=2)
plt.xlabel('Año')
plt.ylabel('Número de artículos')
plt.title('Evolución temporal de las 5 palabras clave más frecuentes')
plt.legend(loc='upper left')
plt.grid(alpha=0.3)
plt.tight_layout()
plt.savefig('grafico_evolucion_keywords.png', dpi=200, bbox_inches='tight')
plt.show()

# =====
# TOPIC MODELLING CON BERTOPIC (PE5)
# =====

import pandas as pd
import matplotlib.pyplot as plt
from bertopic import BERTopic
from sklearn.feature_extraction.text import CountVectorizer

# 1. Cargar datos
df = pd.read_excel("corpus_final_completo.xlsx")

```

```

df_clean = df.dropna(subset=['Abstract']).copy()
abstracts = df_clean['Abstract'].tolist()
years = df_clean['Year'].tolist()
print("Abstracts disponibles:", len(abstracts))

# 2. Configurar el vectorizer para eliminar stopwords
vectorizer = CountVectorizer(stop_words="english", ngram_range=(1, 2))

# 3. Crear el modelo BERTopic con ajustes
topic_model = BERTopic(
    language="english",
    min_topic_size=10,
    n_gram_range=(1, 2),
    vectorizer_model=vectorizer,
    nr_topics=8
)

# 4. Ejecutar
print("Ejecutando BERTopic... (1-3 minutos)")
topics, probs = topic_model.fit_transform(abstracts)
print("¡Listo!")

# 5. Ver temas
print("\n=== TEMAS IDENTIFICADOS ===")
topic_info = topic_model.get_topic_info()
print(topic_info[['Topic', 'Count', 'Name']].to_string())

# 6. Detalle de cada tema (10 palabras)
print("\n=== DETALLE DE CADA TEMA ===")
for topic_id in topic_info['Topic']:
    if topic_id == -1:
        continue
    words = topic_model.get_topic(topic_id)
    top_words = [w[0] for w in words[:10]]

```

```

print(f"\nTopic {topic_id}: {' '.join(top_words)}")

# 7. Asignar topic a cada artículo
df_clean['Topic'] = topics

# 8. Evolución temporal
df_topics = df_clean[df_clean['Topic'] != -1].copy()
evol_topics = df_topics.groupby(['Year', 'Topic']).size().unstack(fill_value=0)

# 9. Gráfico de evolución temporal
top_topics = topic_info[topic_info['Topic'] != -1].nlargest(6, 'Count')['Topic'].tolist()

plt.figure(figsize=(14, 7))
for t in top_topics:
    if t in evol_topics.columns:
        label_words = [w[0] for w in topic_model.get_topic(t)[:3]]
        label = f"Topic {t}: {' '.join(label_words)}"
        plt.plot(evol_topics.index, evol_topics[t], marker='o', label=label, linewidth=2)

plt.xlabel('Año')
plt.ylabel('Número de artículos')
plt.title('Evolución temporal de los principales temas identificados por BERTopic')
plt.legend(loc='upper left', fontsize=8)
plt.grid(alpha=0.3)
plt.tight_layout()
plt.savefig('grafico_evolucion_topics.png', dpi=200, bbox_inches='tight')
plt.show()

# 10. Gráfico de barras
topic_sizes = topic_info[topic_info['Topic'] != -1][['Topic', 'Count']].copy()
topic_sizes['Label'] = topic_sizes['Topic'].apply(
    lambda t: f"Topic {t}: {' '.join([w[0] for w in topic_model.get_topic(t)[:3]])}"
)

```

```

plt.figure(figsize=(10, 6))
plt.barh(topic_sizes['Label'][:, -1],          topic_sizes['Count'][:, -1],          color='steelblue',
edgecolor='black')
plt.xlabel('Número de artículos')
plt.title('Temas identificados por BERTopic y su peso en el corpus')
plt.grid(axis='x', alpha=0.3)
plt.tight_layout()
plt.savefig('grafico_topics_barras.png', dpi=200, bbox_inches='tight')
plt.show()

print("\n=== FIN ===")

print(topic_model.get_topic_info()[['Topic', 'Count', 'Name']].to_string())

# =====
# PARTE 8: Top revistas (PE6)
# =====

# Contar artículos por revista
top_revistas = df['Source title'].value_counts().head(15)
print("\nTop 15 revistas por número de artículos:")
print(top_revistas)

# Gráfico de barras horizontales (mejor para nombres largos)
plt.figure(figsize=(10, 7))
plt.barh(top_revistas.index[:, -1],          top_revistas.values[:, -1],          color='steelblue',
edgecolor='black')
plt.xlabel('Número de artículos')
plt.title("Top 15 revistas con mayor producción sobre IA y banca de inversión")
plt.grid(axis='x', alpha=0.3)
plt.tight_layout()
plt.savefig('grafico_top_revistas.png', dpi=200, bbox_inches='tight')
plt.show()

```