

ANÁLISIS BIBLIOMETRICO DEL USO DE MACHINE LEARNING EN FINANZAS A TRAVÉS DE UN MODELO K-MEANS

Bibliometric analysis of the use of *machine learning* in finance through a k-means model

Jesús Molina-Muñoz

jesus.molina@urosario.edu.co

Universidad del Rosario, Escuela de Administración

No. ORCID: 0000-0002-6578-7516

Colombia

Resumen

El uso de técnicas de machine learning en finanzas se ha convertido en un "hot topic" dentro de la literatura financiera. Lo anterior, se explica por los modernos desarrollos computacionales que han permitido la estimación de modelos, anteriormente desafiantes, con niveles satisfactorios de precisión. En consecuencia, se presenta un aumento significativo en la producción académica que hace uso de este tipo de metodologías en finanzas. El presente trabajo busca explorar las tendencias en términos de producción académica, la concentración por países y autores e identificar la estructura temática dentro de esta área a través de metodologías novedosas como K-means.

Palabras clave

Machine learning, k-means, análisis bibliométrico, gestión financiera

Abstract

The use of machine learning techniques in finance has become a "hot topic" in the financial literature. This is explained by modern computational developments that have allowed the estimation of previously challenging models with satisfactory levels of precision. Consequently, there is a significant increase in the academic production that makes use of this type of methodologies in finance. This work seeks to explore trends in terms of academic production, concentration by countries and authors; and to identify the thematic structure within this area through novel methodologies such as K-means.

Keywords

Machine learning, k-means, bibliometric analysis, financial management.

JEL Code: G,G14, G30

Introducción

En años recientes, el uso de técnicas de inteligencia artificial (IA) ha experimentado un incremento sin precedentes en el campo de la administración. En particular, esta tendencia se ha presentado de manera acentuada en el campo de las finanzas. De acuerdo con datos tomados de la base de datos *web of science* para el año 2001 se encontraba un artículo relacionado con el uso de *machine learning* en finanzas. Para el año 2010 esta cifra se mantenía, mientras que para el año 2020, el número de artículos relacionados con esta temática superaba los treinta.

Lo anterior se explica por dos razones. Primero, los modelos de IA tienden a superar las decisiones tomadas por los humanos en múltiples tareas (Pinar-Saygin, 2000). Segundo, los desarrollos persistentes en computación han permitido que se puedan obtener resultados satisfactorios a través de modelos que anteriormente presentaban un nivel de dificultad en su estimación que los hacía inviables (Dixon *et al.*, 2020).

Estos hechos han ocasionado un interés creciente por parte de practicantes y académicos en el uso de la IA en finanzas. A su vez, esto ha generado un incremento sin precedentes en la producción científica asociada a esta temática (Mora & Norman, 2017). El objetivo de este trabajo es el de identificar y analizar las principales tendencias y la estructura conceptual en materia de producción científica en temas relacionados con el uso de IA en finanzas a través de técnicas novedosas como K-means.

Para cumplir con lo anterior, en el apartado 2 se presenta un breve marco teórico de este tópico, seguido de la metodología usada en la sección 3, para posteriormente pasar a los resultados y las conclusiones en las secciones 4 y 5, respectivamente.

Método

El concepto de *machine learning* se refiere a un conjunto de métodos de *data science*, en los que los modelos pueden aprender de los datos y mejorar su desempeño en el desarrollo de una tarea a través del tiempo (Ghoddussi *et al.* 2019). Los orígenes de este campo de

investigación se pueden rastrear en la década de los cincuenta, donde se presentaron avances relevantes en computación que permitieron la implementación de este tipo de técnicas. Sin embargo, es a partir de los años dos mil cuando se presenta un boom en la literatura científica relacionada con esta temática. Lo anterior, debido principalmente a los avances computacionales que se presentaron en las últimas décadas. Las ventajas derivadas de los resultados obtenidos a partir de este tipo de técnicas han permitido que tengan gran acogida en diversas disciplinas. Aplicaciones de este tipo de modelos pueden ser encontrados en el ámbito científico, tecnológico y comercial. Lo anterior ha posibilitado el surgimiento de nueva evidencia de aplicaciones exitosas de *machine learning* en procesos de toma de decisiones en los campos de la salud, manufactura, educación, modelación financiera y *management*. (Jordan y Mitchell, 2015).

La IA ha recibido gran acogida dentro de las finanzas como campo de investigación debido al amplio uso de herramientas computacionales dentro de este campo y a la relativa abundancia de los datos utilizados en esta área. Los primeros trabajos seminales en este campo se pueden rastrear hasta la década de los noventa con propuestas relacionadas con el uso de redes neuronales en la toma de decisiones financieras y administración de riesgos financieros (Hawley *et al.* 1990; Altman *et al.* 1994; Varetto, 1998).

En años recientes, el uso del *machine learning* en las finanzas se ha acentuado debido a los avances en métodos computacionales (Dixon *et al.* 2020) y la disponibilidad de nuevas fuentes de información diferentes a las tradicionalmente usadas (de Prado, 2019). En particular, una gran parte de los trabajos de *machine learning* en finanzas se podrían clasificar en áreas relacionadas con técnicas de pronóstico de precios, análisis de mercados financieros, pronóstico de riesgo, perspectivas financieras y finanzas corporativas (Aziz *et al.* 2019).

Para el desarrollo del presente trabajo se siguieron los siguientes pasos. Primero, se descargaron los datos de las bases Web of Science y Scopus consideradas las más

importantes en materia de producción científica. El periodo seleccionado fue el comprendido entre 1985 y 2020. El total de número de artículos encontrados en la primera base de datos fue de 129. Para Scopus, el total de artículos encontrados fue de 705. La búsqueda se realizó a través de los términos consignadas en la tabla 1.

Machine learning finance
Artificial intelligence finance
Neural networks finance
Big data finance
Decision trees finance
Support vector machine finance
Supervised learning finance
Deep learning finance
Unsupervised learning finance
Ensemble method finance
Genetic algorithm finance
Particle swarm optimization finance

Tabla 1. Palabras clave búsqueda en Web of Science y Scimago

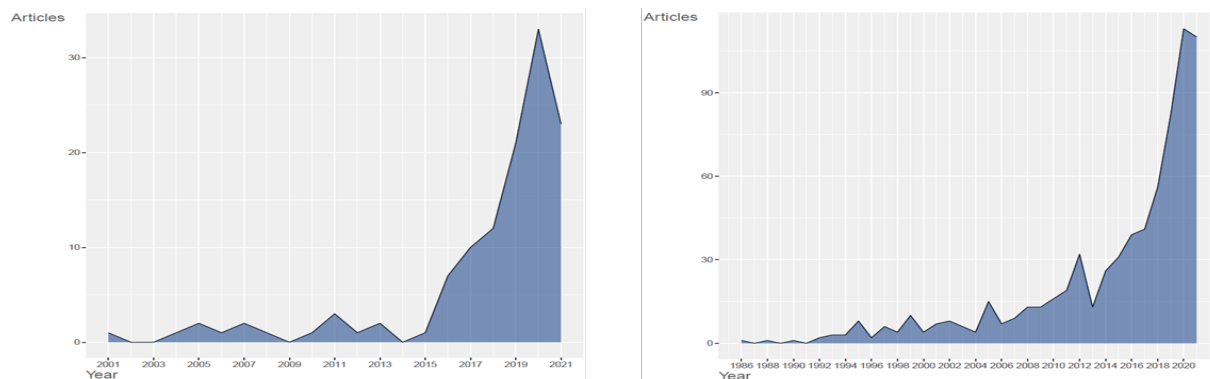
Como segundo paso, se realizó un análisis estadístico en términos de número de productos por país, autor y número de citas para los productos obtenidos a partir del desarrollo del primer procedimiento. Este análisis se hizo a través del paquete Bibliometrix en el software R (Aria y Cuccurullo, 2017). Posteriormente, se condujo un análisis de redes para identificar la concentración entre los autores de los artículos y sus respectivos países. Finalmente, se llevó a cabo un análisis de K-means² para hacer un análisis de la estructura conceptual del uso de *machine learning* en finanzas.

² Para revisar los detalles acerca del método k-means es posible consultar Xiong y Chen (2008).

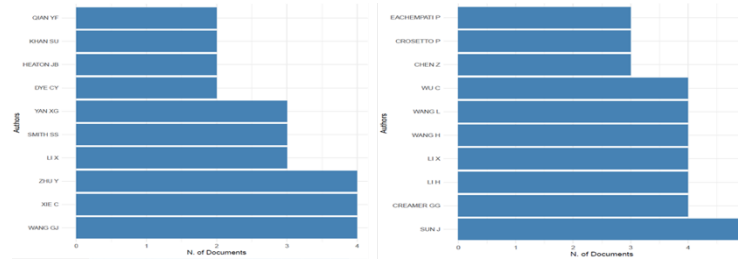
Resultados

En términos de producción (gráfica 1), el número de productos relacionados con *machine learning* en finanzas ha aumentado de manera sostenida en los últimos 20 años. En particular se observa un boom en el número de artículos científicos a partir del año 2015 tanto para WoS como para Scopus.

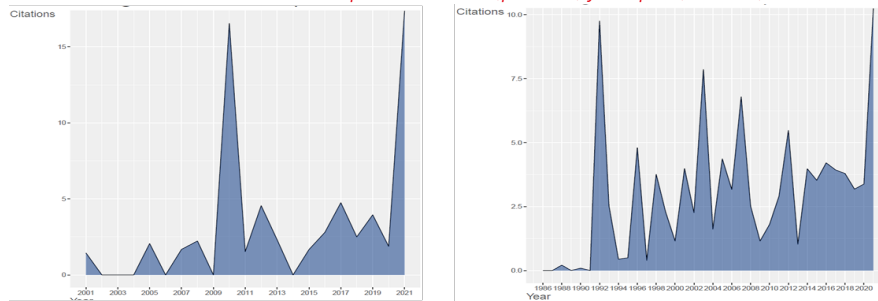
Al revisar el número de productos por autor, se observa coincidencia para autores destacados en este campo como Wang, Li X. y Zhu (gráfica 2) para las dos bases de datos. Este dato está acompañado de un aumento sostenido en el número de citas promedio por año (gráfico 3) que, si bien presenta diferentes picos, presenta su valor más alto en los años 2020 y 2021 en las dos bases de datos. Este hecho confirma la relevancia actual que presenta esta temática en la producción científica en finanzas.



Gráfica 1. Número de productos por año WoS (panel A) y Scopus (Panel B)

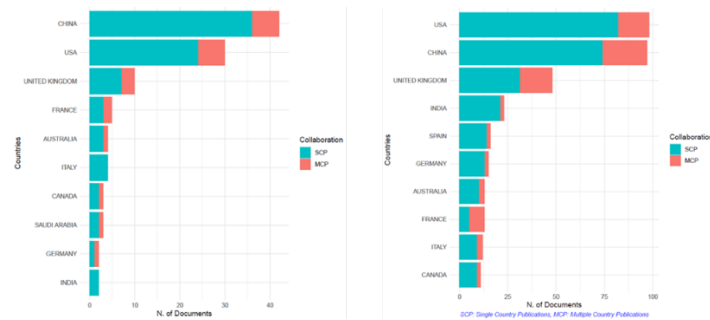


Gráfica 2. Producción por autor WoS (panel A) y Scopus (Panel B)

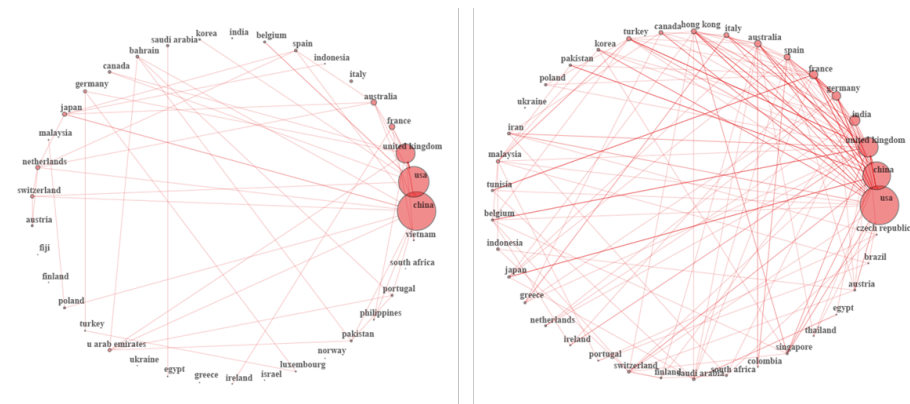


Gráfica 3. Promedio de citaciones por año WoS (panel A) y Scopus (Panel B)

Al revisar la producción por países se observa una tendencia bastante clara para las dos bases de datos. La mayor parte de la producción está concentrada en U.S.A., China y Reino Unido. Llama la atención que, al construir una red de acuerdo con la afiliación de los autores por país, para analizar las colaboraciones entre académicos en esta área, la mayor concentración de colaboraciones se presenta en los países mencionados (gráfica 4) seguidas de contribuciones menores por parte de autores de India, Francia y Alemania. Este hecho evidencia una gran concentración de la producción académica en este campo de investigación.



Gráfica 4. Producción por país WoS (panel A) y Scopus (Panel B)



Gráfica 5. Red colaboración entre países WoS (panel A) y Scopus (Panel B)

Al analizar la estructura conceptual a través del método de *clustering K-means* (gráfico 6), se presenta una clara tendencia en términos de concentración para artículos relacionados con inversiones, mercados financieros y *fintech* a través del uso de técnicas relacionadas con *Deep learning* y redes neuronales. Otro grupo conceptual se observa para métodos de *machine learning* como análisis discriminante, árboles de decisión y análisis de regresión para temas relacionados con administración de riesgos, decisiones financieras y variables económicas.

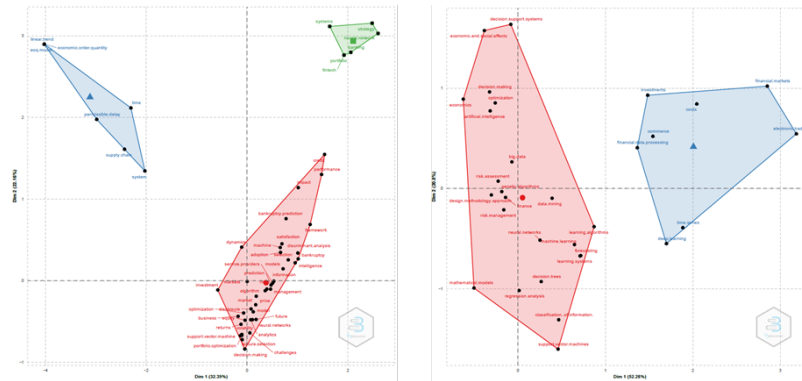


Gráfico 6. Estructura conceptual (clusters por k-means) WoS (panel A) y Scopus (Panel B)

Discusión y conclusión

El análisis realizado permite concluir que en los últimos años se ha presentado un incremento significativo en el número de productos relacionados con *machine learning* en finanzas. Esta tendencia experimenta un aumento sin precedentes a partir del año 2015. Si bien no se presenta un alto grado de concentración por autores, si se evidencia una concentración importante en términos de número de productos y colaboraciones para U.S.A, China y Reino Unido. Al revisar la estructura conceptual de los productos analizados, se observa un uso relevante de metodologías como *Deep learning* y redes neuronales para el análisis de inversiones y mercados financieros y otras técnicas de IA para toma de decisiones y riesgos financieros.

Lo anterior sugiere un mayor esfuerzo por parte de otros países en términos de producción científica en esta área de investigación, así como la necesidad de incrementar la investigación en temáticas relevantes dentro del campo financiero como las finanzas corporativas, el análisis de mercados emergentes y la predicción y transmisión de volatilidad que se encuentran ausentes en la evidencia obtenida del trabajo realizado.

Referencias bibliográficas

- Aziz, S., Dowling, M., Hammami, H., & Piepenbrink, A. (2019). *Machine learning* in finance: A topic modeling approach. *European Financial Management*.
- Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of banking & finance*, 18(3), 505-529.
- Aria, M. & Cuccurullo, C. (2017) bibliometrix: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics*, 11(4), pp 959-975, Elsevier.
- Aziz, S., Dowling, M., Hammami, H., & Piepenbrink, A. (2019). *Machine learning* in finance: A topic modeling approach. *European Financial Management*.
- Dixon, M. F., Halperin, I., & Bilokon, P. (2020). *Machine learning* in Finance. Springer International Publishing.
- Ghoddussi H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, 81, 709-727.
- Hawley, D. D., Johnson, J. D., & Raina, D. (1990). Artificial neural systems: A new tool for financial decision making. *Financial Analysts Journal*, 46(6), 63-72.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- López de Prado, M. (2019). Beyond econometrics: A roadmap towards financial machine learning. Available at SSRN 3365282.
- Pinar Saygin, A., Cicekli, I., & Akman, V. (2000, November). Turing test: 50 years later. *Minds Mach.*, 10(4), 463-518.
- Mora Ramírez, A. J., & Norman Acevedo, E. (2017). Revisión sistemática de literatura sobre la internacionalización de la empresa PYME en Colombia. *Espacios*, 38, 39.
- Varetto, F. (1998). Genetic algorithms applications in the analysis of insolvency risk. *Journal of Banking & Finance*, 22(10-11), 1421-1439.
- Xiong, H., Wu, J., & Chen, J. (2008). K-means clustering versus validation measures: a data-distribution perspective. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(2), 318-331.