

AI and ML in accounting and finance: A bibliometric review

Alessandra Belfiore¹, Agostino Gnasso², Corrado Cuccurullo³, Massimo Aria⁴

¹Università degli Studi della Campania "Luigi Vanvitelli" – alessandra.belfiore@unicampania.it

²Università degli Studi di Napoli "Federico II" – agostino.gnasso@unina.it

³Università degli Studi della Campania "Luigi Vanvitelli" – corrado.cuccurullo@unicampania.it

⁴Università degli Studi di Napoli "Federico II" – aria@unina.it

Abstract

Artificial intelligence (AI) and machine learning (ML) are two related technologies in accounting and finance studies. This study maps the conceptual structure of AI and ML research with the aim of contributing to a better understanding of this research stream. A bibliometric analysis of 3,836 documents on ai and ML retrieved from the Web of Science database is conducted. The analysis of descriptive performance indicators identifies the main traits of the scientific debate about AI and ML in terms of publications, productive countries and sources. To map the conceptual structure of the dataset, the study performs a thematic evolution. The results highlight the growing academic interest in the research topic, especially in the past few years. The results of this study may provide scholars with a better understanding of AI and ML research in accounting and finance. This paper contributes to the field by providing an examination of the current state of the art of AI e ML research and identifying possible future research directions.

Keywords: Accounting, Finance, Artificial Intelligence, Machine Learning, Conceptual structure, Journal ranking

1. Introduction

The growing success of artificial intelligence (AI) and machine learning (ML) in the application and research domains of accounting and finance is revolutionizing industries and corporations (Li and Tang, 2020). Hedge funds, retail banks, and FinTech, supported by massive growth in computing power and storage, are investing heavily in the acquisition of skills in data science and ML (Wall, 2018). Furthermore, together with the continuous search for innovation with the use of big data, there is an ever increasing need to observe regulatory systems. For example global financial crisis in 2007 led to major structural changes in the financial industry, pushing tight regulation on the topic of big data. De Prado (2019) argues that impede to use classical econometric linear models are unable to predict due to different types of data, such as using data such as voice recordings, news articles, social media posts media such as Tweets, photos and satellite images, called alternative data. Hence, ML models allow to provide computational power and flexibility to run complex models in high-dimensional data environments. Moreover, the scientific community searching for more accuracy and interpretability of these ML models, are starting to use models that exploit the internal relationships between features in order to

perform interpretative analyses, predictive analyses and casual inference (Aria, Gnasso, Cuccurullo, 2021).

This study aims to highlight how AI technology and ML models have influenced the accounting and finance industries in the last decade. The contribution to the existing literature can be identified from several points of view. In particular, this study examines the related AI-Accounting and Finance literature, published between 2010 and 2021, and summarizes its focused areas, allowing to highlight the trend and nature of the contents found in the current literature. Secondly, this work shows the current application of AI and ML in the fields of accounting and finance, which aims to highlight the relevant areas, and can certainly serve as a reference point for future studies in this domain.

2. Materials and Methods

2.1. Data collection

To retrieve the bibliometric data, we used the Web of Science (WoS), querying SCI-EXPANDED (Science Citation Index Expanded) and SSCI (Social Science Citation Index) databases. These two databases appear to be the most used in the social sciences (Zupic, Čater, 2015) and have standardized fields that require little data cleaning. We used the following search terms: "Artificial Intelligence", "Machine Learning", "Deep Learning", "Data Science" in the title, abstract, and keywords fields over the period 1985-2022, focusing only on English documents. Few articles had been published before 2010, while in 2022 publications are still in progress when the data was downloaded. So, we focused on the 2010- 2021 period. Also, there were various document types, but we decided to only consider the articles, proceedings paper, review. Finally, we focused only on two subject area: "business", "finance" and "management". This query returned 3,836 publications.

2.2. Data analysis

We analyzed the whole dataset with the Bibliometrix R package (Aria, Cuccurullo, 2017). It is a unique tool according to a logical bibliometric workflow and incorporates a wide variety of different analyses. This tool is the only one that allows you to control and follow the entire bibliometric analysis workflow from importing bibliographic data from various databases, including the Clarivate Analytics Web of Science, to performing bibliometric analysis and data visualization. It is open-source and returns statistical analysis that other bibliometric tools do not provide.

After some descriptive statistics, we extracted the conceptual knowledge structure of AI and ML research in accounting and finance. This structure is used to understand the topics covered by a research field to define the most important and recent issues (the so-called research front) and to study the evolution of concepts over time (Börner et al., 2003). The methodological foundation of the conceptual structure is the idea that the co-occurrence of key terms describes the content of the documents in a dataset (Callon et al., 1991, Cobo et al., 2011). We used the KeywordsPlus to identify thematic clusters.

By applying the Louvain clustering algorithm on the co-word network, we highlighted the different themes of a given domain. This thematic network is drawn on a two-dimensional matrix, where the axes are function of the Callon centrality and the Callon density of the thematic map. The centrality can be read as the relevance of the theme in the whole research field. Density is intended as a measure of the degree of development of the theme.

The two-dimensional matrix defines four quadrants related to different themes (Callon et al., 1991; Coulter et al., 1998; He, 1999; Cahlik, 2000; Aria et al., 2020): (i) motor themes (upper right quadrant) which are very relevant and very well developed; (ii) basic themes (lower right quadrant) that are very topical but not well developed; (iii) emerging or declining themes (lower left quadrant) not very relevant and developed; (iv) niche themes (upper left quadrant) not very relevant but very well developed.

Each theme represented in the bi-dimensional matrix is a co-word network cluster. The names in the bubble are the words with the highest occurrence value. The size of the bubble is proportional to the occurrences of the words in the cluster.

For a longitudinal analysis, we articulated the time span into 3 time slices. Looking at the distribution of publications per year, we have decided to divide our collection into 3 equal time spans (2010-2013; 2014-2017; 2018-2021), with 2 cutting points (2013 and 2017).

3. Findings

3.1. Descriptive findings

Table 1 shows significant information about our collection. The number of documents on AI and ML is growing exponentially. In the period from 2010 to 2021, the annual growth rate is 16.71%. This exponential growth rate indicates that academic production is intensifying. This is probably due to the fact that the topic of AI and ML in accounting and finance is global. The analysis of national scientific production showed that the number of documents in which at least one author comes from a particular country varies from 1 for Iraq, Montenegro, Philippines to 3,730 for the USA (Figure 1). The most productive countries for publications related to AI and ML in accounting and finance are the United States of America (USA) (n=3,730; 27%), followed by China (n=1,677; 12%), the United Kingdom (UK) (n=1,247; 9%), Germany (n=795; 6%).

The 3,836 documents were written by 9,080 authors on 299 different journals that are all non-leading journals (ABS 2018 ranks < 3). This reflects the ABS 2018 ranking in which 70% (62/89) of accounting journals and 66% (73/110) of finance journals are ranked < 3.

As shown in Figure 2, the most relevant source is Technological Forecasting and Social Change with 258 publications (7%). Approximately, 4% of the documents are published by European Journal of Operational Research (n=165) and Research Policy (n=148). Additional sources for AI and ML include the following journals: Journal of Business Research (n=112; 3%), Science and Public Policy (n=84; 2%), Journal of Technology Transfer (n=70; 2%), and International Journal of Forecasting (n=67; 2%).

All the publications are associated with 11,082 authors' keywords and 5,915 keywords assigned by the WOS machine learning algorithm.

MAIN INFORMATION ABOUT DATA	
Timespan	2010:2021
Sources (Journals, Books, etc)	299
Documents	3836
Average citations per documents	22,19
Average citations per year per doc	3,577
References	181027
DOCUMENT TYPES	
article	3578
article; book chapter	7
article; proceedings paper	45
review	206
DOCUMENT CONTENTS	
Keywords Plus (ID)	5915
Author's Keywords (DE)	11082
AUTHORS	
Authors	9080
Author Appearances	11366
Authors of single-authored docs	467
AUTHORS COLLABORATION	
Single-authored docs	486
Documents per Author	0,422
Co-Authors per Doc	2,96
International co-authorships %	38,35

Table 1 Main information about data

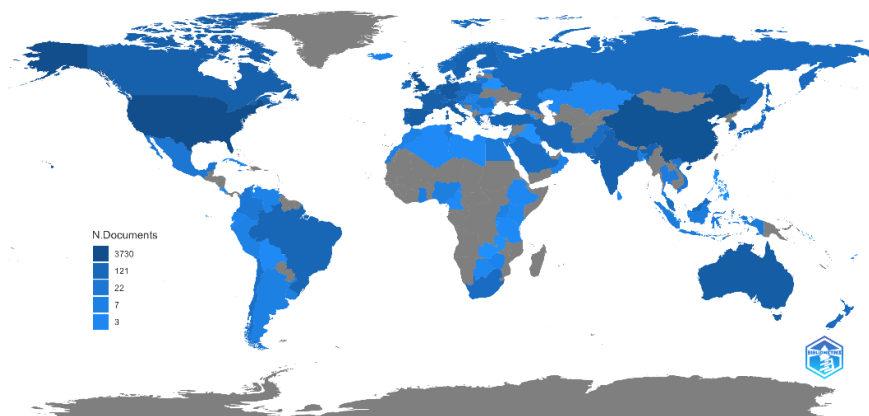


Figure 1 Country scientific production

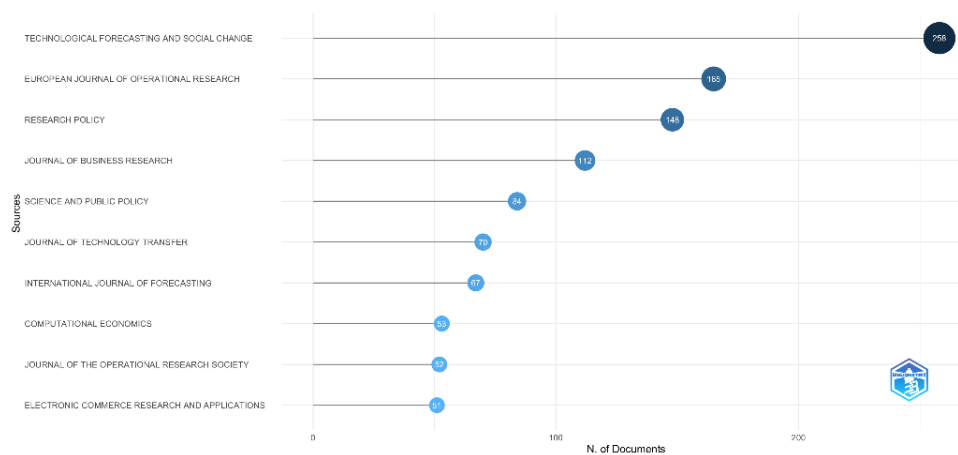


Figure 2 Most relevant sources

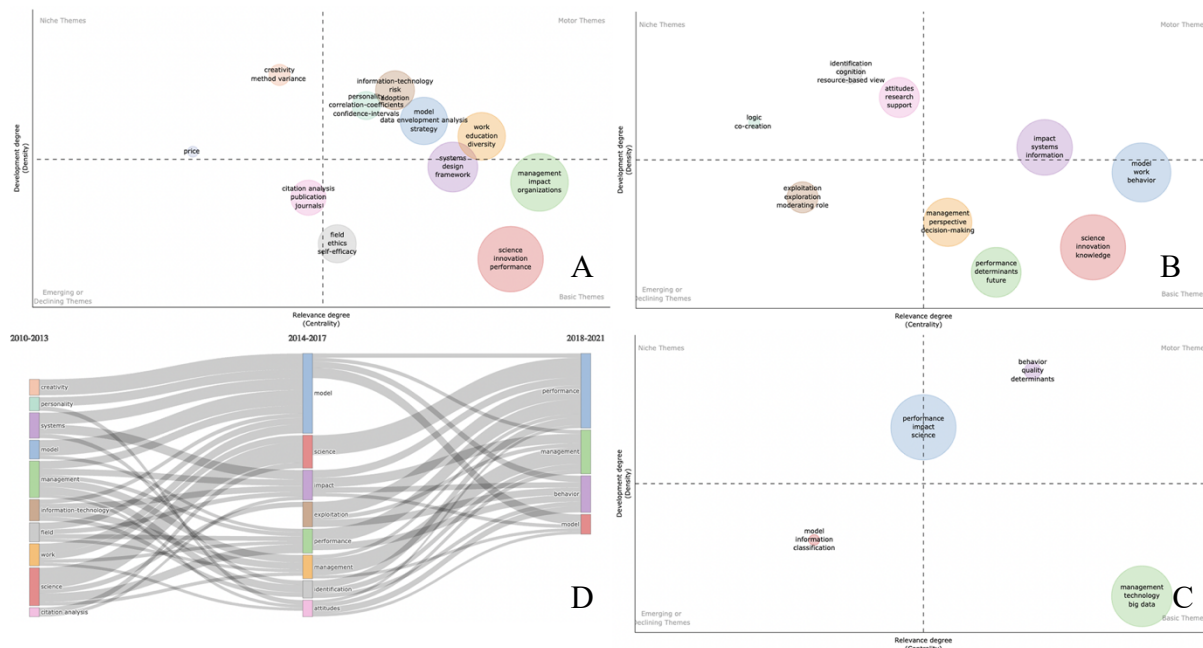


Figure 3 A, time slice 2010-2013; B, time slice 2014-2017; and C, time slice 2018-2021; D, Thematic evolution of Keywords Plus in field of research on AI and ML (2010-2021).

3.2. Conceptual structures evolution

Figure 3 reports the results of the analysis of the thematic evolution of publications in the three periods considered (2010-13; 2014-17; 2018-21). Looking at the Callon centrality and density, we analyze the motor themes, the basic themes, the emerging or declining themes, and the niche themes. Comparing the thematic map of the three periods considered, we analyze and trace the evolution of the themes (in terms of trajectory on the theme). In the first period 2010-13, the research identified 10 thematic areas (Figure 3B): "information-technology", "strategy", "education", "personality" as a motor theme, "systems", "management", "innovation" and "ethics" as a base theme, "citation analysis" as an emerging theme and "creativity" as a niche theme. In the second period 2014-17 (Figure 3C), all the motor themes become basic or niche themes. Finally, in the third period 2018-21, we have a concentration in 4 thematic areas (Figure 3D). The most relevant and developed topics in the research on AI and ML in accounting and finance is "quality" and "impact" themes. Instead, "management" is a basic theme, while "information" as an emerging or declining theme.

4. Conclusion

This article examined the existing literature on AI and ML to study the conceptual framework of research in accounting and finance. We performed a bibliometric analysis divided into two steps: first, we have made a descriptive analysis; second, we have identified the conceptual knowledge structure. Our findings highlight a growing and global academic interest in this research topic. Recommendations on what potential journals could be selected for publication. We conclude that no studies have been published in top journals. The analysis of the thematic evolution has allowed us to understand the research themes and/or thematic changes that have accompanied the growth of scientific production on AI and ML in accounting and finance. Keyword Plus analysis shows how AI and ML research in accounting and finance is focusing on quality, operations, impact and performance. Our study has both theoretical and practical implications. Firstly, this study contributes to the existing literature on the topic of AI and ML

in accounting and finance. This analysis provides a starting point for scholars interested in studying this research topic in several ways. The study presents the longitudinal evolution of AI and ML research, revealing the most important themes related to AI and ML of each period. The conceptual structure shows that criticalities have already been developed, while theoretical and empirical areas remain to be explored. Secondly, our research highlights how the term “impact” has become relevant in the conceptual map. This may also present limits from which to define potential directions for future research.

This study focused on international articles, paper proceedings, and peer-reviewed publications only in journals indexed on WoS. As a result, different types of data on the subject (such as patents) were not included in our analysis. We have limited our investigation to the Plus keywords only. It would be interesting to develop further conceptual analyzes (such as LDA).

References

- Aria, M., & Cuccurullo, C. (2017). *bibliometrix: An r-tool for comprehensive science mapping analysis*. *Journal of informetrics*, 11, 959–975.
- Aria, M., Cuccurullo, C., & Gnasso, A. (2021). A comparison among interpretative proposals for random forests. *Machine Learning with Applications*, 6, 100094.
- Aria, M., Misuraca, M., & Spano, M. (2020). Mapping the evolution of social research and data science on 30 years of social indicators research. *Social indicators research*, 149, 803–831.
- Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual review of information science and technology*, 37, 179–255.
- Cahlik, T. (2000). Comparison of the maps of science. *Scientometrics*, 49, 373–387.
- Callon, M., Courtial, J.-P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, 22, 155–205.
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for information Science and Technology*, 62, 1382–1402.
- Coulter, N., Monarch, I., & Konda, S. (1998). Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science*, 49, 1206–1223.
- He, Q. (1999). Knowledge discovery through co-word analysis.
- Harvey, C., Kelly, A., Morris, H. and M. Rowlinson (2010), *Academic Journal Quality Guide, Version 4*, London: The Association of Business Schools.
- Li, X., & Tang, P. (2020). Stock index prediction based on wavelet transform and fcd-mlgru. *Journal of Forecasting*, 39, 1229–1237.
- López de Prado, M. (2019). *Beyond econometrics: A roadmap towards financial machine learning*. Available at SSRN 3365282, .
- Wall, L. D. (2018). Some financial regulatory implications of artificial intelligence. *Journal of Economics and Business*, 100, 55–63.
- Zupic, I., & Cater, T. (2015). Bibliometric methods in management and organization. *Organizational research methods*, 18, 429–472.