International Journal of Informatics and Computation (IJICOM)
Vol. 7, No. 2, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Explainable Artificial Intelligence for Financial Services: A Bibliometric Review

Danang Junaedi[1], Suyanto[2], Mira Kania Sabariah[2], Rio Guntur Utomo[2]

Abstract

Artificial intelligence has become an essential technology in the banking, financial services, and insurance (BFSI) sector, supporting tasks such as risk assessment, fraud detection, and financial forecasting. However, the increasing use of complex "black-box" models, including deep learning and generative AI, raises serious concerns related to transparency, trust, and regulatory compliance. This study aims to provide a clear overview of the research landscape on Explainable Artificial Intelligence (XAI) in financial services by identifying key trends, commonly used methods, and existing research gaps. We analyzed 580 publications indexed in the Scopus database from 2018 to 2024 using bibliometric techniques, supported by VOSviewer and Latent Dirichlet Allocation for thematic analysis. The results reveal a rapid growth in XAI-related publications, particularly after 2021, with post-hoc explanation methods such as SHAP and LIME being the most widely adopted. At the same time, our findings indicate that explainability remains limited for emerging generative models, including GANs and Transformer-based architectures, especially in applications like fraud detection and financial prediction. Overall, this paper provides practical insights and a strategic reference for researchers and practitioners seeking to align advanced AI models with the transparency and accountability requirements of the financial sector.

Keywords:
Review, Bibliometric, Analysis, Finance, XAI

1. Introduction

The rapid adoption of artificial intelligence (AI) across financial services fundamentally transforms decision-making processes in areas such as credit scoring, fraud detection, risk management, and financial forecasting. Market reports indicate sustained and accelerating investment in AI-driven solutions, driven by their ability to process large-scale financial data and improve operational efficiency and predictive accuracy. However, the increasing reliance on complex machine learning and deep learning models introduces significant transparency and accountability concerns, particularly when these models influence high-stakes financial decisions. Financial institutions face growing pressure from regulators, customers, and internal governance bodies to justify automated decisions in a manner that is understandable and auditable, positioning explainability as a critical requirement rather than an optional feature [1], [5].

Explainable Artificial Intelligence (XAI) emerges as a response to the opacity of black-box models by enabling human-understandable explanations of model behavior, predictions, and decision logic. Foundational studies in XAI establish conceptual frameworks, taxonomies, and methodological classifications that distinguish between model-intrinsic and post-hoc explainability, as well as between global and local

Corresponding Author: Danang Junaedi(danangjunaedi@telkomuniversity.ac.id)
1 Danang Junaedi, School of Computing Telkom University Bandung, danangjunaedi@telkomuniversity.ac.id
2 Suyanto, School of Computing Telkom University Bandung, suyanto@telkomuniversity.ac.id
3 Mira Kania Sabariah, School of Computing Telkom University Bandung, mirakania@telkomuniversity.ac.id
4 Rio Guntur Utomo, School of Computing Telkom University Bandung, riogunturutomo@telkomuniversity.ac.id

explanations. Despite these advances, applying XAI in real-world financial contexts remains challenging due to the complexity of financial data, the nonlinear nature of modern models, and the trade-off between predictive performance and interpretability. These challenges motivate a systematic examination of how XAI concepts translate into practical financial applications [3], [2].

The financial sector presents unique explainability demands because automated decisions directly affect individuals' access to credit, investment opportunities, and financial security. Regulatory frameworks increasingly emphasize transparency, fairness, and accountability, compelling financial institutions to adopt explainable models that align with legal and ethical requirements. Empirical studies involving banks and supervisory authorities highlight a disconnect between technical XAI methods and regulatory expectations, particularly regarding explanation fidelity, consistency, and usability for non-technical stakeholders. This gap underscores the need for a structured synthesis of existing research to understand how XAI is operationalized in financial services and where current approaches fall short [4], [5].

A growing body of literature focuses on XAI in financial time series forecasting, including stock prices, exchange rates, and market risk indicators. These studies demonstrate that explainability techniques such as SHAP, LIME, and attention mechanisms can enhance trust and insight without substantially degrading predictive performance. However, findings also reveal inconsistencies in evaluation practices, limited comparability across studies, and insufficient attention to explanation stability over time. As financial markets exhibit non-stationary behavior, explainability methods must address both temporal dynamics and model adaptability, which remains an open research challenge [6], [23].

Credit scoring represents one of the most extensively studied application domains for XAI in finance due to its regulatory sensitivity and societal impact. Prior research integrates explainable methods with traditional and deep learning-based credit models to reveal feature importance, decision boundaries, and individual-level explanations. While these approaches improve transparency, studies also report increased computational cost, reduced scalability, and potential risks of explanation misuse or oversimplification. Furthermore, the lack of standardized explainability benchmarks complicates the assessment of competing XAI techniques in credit-related tasks [15], [17], [26].

Beyond credit risk, XAI research extends to fraud detection, financial risk monitoring, and algorithmic investment strategies. In these domains, explainability supports anomaly investigation, regulatory audits, and strategic decision-making. However, existing studies often focus on isolated use cases or specific algorithms, limiting their generalizability. Additionally, the interaction between explainability, data privacy, and security introduces new constraints, especially when explanations may inadvertently expose sensitive information. These multidimensional challenges necessitate a holistic overview of XAI adoption across financial services [19], [25], [28].

Despite the expanding literature, prior reviews predominantly adopt narrative or domain-specific perspectives, offering limited quantitative insight into research trends, influential contributions, and collaboration networks. Bibliometric analysis provides a systematic and reproducible approach to mapping the intellectual structure of a research field by analyzing publication patterns, citation networks, thematic evolution, and methodological convergence. Established guidelines for systematic mapping and bibliometric studies emphasize their suitability for identifying research gaps, maturity levels, and emerging topics, yet such approaches remain underutilized in XAI-focused financial research [10], [11].

In response to these gaps, this study conducts a comprehensive bibliometric review of Explainable Artificial Intelligence in financial services to analyze publication growth, dominant application domains, methodological trends, and thematic clusters. By synthesizing insights from a diverse body of interdisciplinary research, this paper aims to clarify how XAI evolves within finance, identify unresolved challenges, and highlight promising research directions. This contribution supports researchers, practitioners, and policymakers in advancing transparent, trustworthy, and responsible AI systems for financial decision-making [2], [6], [27].

2. Related Work

Several studies concentrated on the trade-off between predictive performance and interpretability in financial AI systems. Dessain *et al.* analyzed the cost of explainability in credit scoring models and demonstrated that enforcing explainability constraints often reduced model accuracy, especially when replacing complex nonlinear learners with interpretable linear approximations [13]. The authors quantified this trade-off and showed that explainability introduced measurable economic and operational costs. However, the study focused primarily on static credit scoring datasets and did not evaluate adaptive or sequential financial models, limiting its applicability to dynamic financial environments.

Pamuk and Schumann investigated machine learning models for forecasting corporate credit ratings using annual financial statements [14]. Their study demonstrated that ensemble and deep learning models outperformed traditional statistical approaches, but the authors also noted that these models lacked transparency. Although the paper acknowledged explainability as a critical challenge, it did not integrate XAI techniques directly into the modeling pipeline, leaving unresolved questions about how explanations could support regulatory and managerial decision-making.

Several researchers explored hybrid and fuzzy-based explainable models to address the black-box nature of deep learning. Chimatapu *et al.* proposed hybrid deep learning and type-2 fuzzy logic systems that preserved high predictive performance while providing linguistic rule-based explanations [16]. Their work illustrated how fuzzy logic enhanced interpretability in complex models. However, the approach increased computational complexity and required expert knowledge to design fuzzy rules, which constrained scalability in large financial datasets.

Local explanation techniques also received significant attention in the literature. La Gatta *et al.* introduced CASTLE, a cluster-aided explanation framework that improved the stability and locality of explanations for black-box classifiers [18]. Their experiments demonstrated improved consistency compared to traditional local explanation methods. Despite these strengths, the method required additional preprocessing steps and clustering assumptions, which could affect robustness in high-dimensional financial data.

Case-based reasoning emerged as another explainability-driven approach in financial risk analysis. Li *et al.* developed a data-driven explainable case-based reasoning framework for financial risk detection [19]. Their system justified predictions by retrieving similar historical cases, which enhanced user trust and interpretability. However, the reliance on historical similarity limited the model's ability to generalize to rare or previously unseen financial events. Peer-to-peer lending and alternative credit markets also motivated explainable scoring models. Lyócsa *et al.* compared default-based and profit-based credit scoring systems and demonstrated that explainable scoring criteria improved transparency for lenders and borrowers [20]. The authors highlighted that profit-optimized models often conflicted with interpretability goals. Nevertheless, the study focused on traditional feature-based models and did not explore deep learning explainability in these markets.

Explainability in financial time series forecasting attracted increasing attention. Argotty-Erazo *et al.* proposed an interpretable linear-model-based methodology for predicting currency exchange rate movements [21]. Their approach provided clear economic interpretations of model coefficients but sacrificed predictive power compared to nonlinear models. Similarly, Ghosh *et al.* integrated nonlinear feature transformations with advanced AI to predict stock market movements influenced by media sentiment [22]. While their framework improved accuracy, explainability remained limited and largely post-hoc.

Several authors investigated the integration of XAI into deep learning forecasting systems. Buğra *et al.* demonstrated that incorporating explainable AI techniques enhanced machine learning predictions in financial time series tasks without severely degrading performance [23]. Sisodia and Khare further compared deep learning models combined with interpretability techniques and showed that attention mechanisms and SHAP-based explanations improved transparency [24]. However, both studies highlighted the absence of standardized metrics for evaluating explanation quality.

Fraud detection and security-sensitive applications further reinforced the importance of explainability. Amirineni proposed explainable AI models for fraud detection in banking ecosystems and demonstrated improved accountability and auditability [25]. Talaat *et al.* integrated explainable AI with deep learning for credit card default prediction and showed that explanation-aware models increased stakeholder trust [26]. Despite their success, these approaches often required extensive feature engineering and domain expertise.

Recent studies emphasized the need for standardized and human-centered explainability frameworks. Goram proposed interactive AI transparency labels to communicate model behavior to non-technical stakeholders [27]. Chakkappan *et al.* examined explainable AI and big data analytics for financial security and privacy risks, emphasizing regulatory compliance and explainability-driven governance [28]. These works highlighted that explainability extended beyond algorithms and required system-level design, yet they did not fully address how to harmonize interpretability with high-frequency, data-intensive financial applications.

Finally, emerging research explored explainability in hybrid forecasting and investment systems. Studies combining LSTM, GRU, ARIMA, and CNN architectures demonstrated strong predictive performance in financial markets but reinforced the opacity of deep models [29]–[31]. Recent surveys and conceptual frameworks called for multi-dimensional evaluation of explainability in credit risk and fraud detection [32]–[34]. Collectively, these works underscored persistent gaps in standardized evaluation, scalability, and domain-specific explanation design, motivating the need for systematic literature analysis on explainable artificial intelligence in financial services.

3. Research Method

Adopting established systematic mapping and bibliometric guidelines [10], [11], this paper utilized a five-stage research methodology to ensure rigor, transparency, and reproducibility in analyzing explainable artificial intelligence research within financial services. We structured the methodology to progressively refine the scope, collect relevant literature, filter high-quality studies, extract meaningful insights, and visualize research trends. This structured approach enabled us to comprehensively capture the evolution, intellectual structure, and thematic patterns of explainable artificial intelligence in the financial domain.

In the planning stage, we defined the research scope, formulated research questions, and established search strategies along with selection criteria. This paper focused on explainable artificial intelligence in financial services, including aspects such as transparency, interpretability, trust, and responsible artificial intelligence. We derived the research questions presented in Table 1 to guide the bibliometric investigation. These questions examined publication growth, thematic scope, topic evolution, potential interrelated research areas, influential authors, and dominant publication venues. By clearly defining these objectives, we ensured consistency between the research design and the analytical outcomes. In this paper, we utilize some research questions in Table 1.

Table 1. Bibliometric Analysis Questions List.

No	Question	Objective
1	From 2018 through 2024, how has the publication population influenced research development?	Examining the number of publications in the field of XAI services over the past five years can shed light on its research development.
2	Just what does the scope of the study encompass?	To find topics that are relevant to the research and keywords that appear together
3	How has the field of study evolved?	Ten sub-topics are identified to ascertain whether the number of publications according to yearly categories continues to rise. This is done to comprehend the study patterns derived from the specified topic coverage using extracted co-occurring keywords.
4	What are the potential interrelated areas of research that could yield fruitful results?	Finding related subjects by analyzing the topic's co-occurring keywords within the specified scope and the trend over the past five years
5	To what extent have certain writers significantly impacted the field of study?	To determine which writers have had the most impact on the field of study
6	In light of the study's parameters, which forms of publication are most significant in light of the study's parameters?	To determine which publishing genres and publishers have the most impact within the parameters of the research

To operationalize the research questions in Table 1, we developed a comprehensive search strategy. We utilized a wide range of keywords and synonyms related to explainable artificial intelligence, such as interpretable artificial intelligence, transparent artificial intelligence, understandable artificial intelligence, and responsible artificial intelligence, combined with finance-related terms using wildcard expressions such as financ*. We applied Boolean operators to capture diverse terminologies used across disciplines. Furthermore, we restricted the publication timeframe to studies published between 2018 and 2024 to reflect the rapid growth of explainable artificial intelligence research and ensure the relevance of the collected literature.

In the search execution stage, we conducted the literature search exclusively using the Scopus database, which we selected due to its broad coverage of peer-reviewed journals and conference proceedings across multiple disciplines. Scopus provided a reliable and comprehensive source for bibliometric analysis. Using the predefined search query, we retrieved 868 documents that met the initial inclusion criteria. At this stage, we emphasized maximizing coverage to ensure that potentially relevant studies were not excluded prematurely.

During the primary study selection stage, we refined the dataset by applying stricter inclusion and exclusion criteria. We limited the document types to journal articles and conference papers to maintain academic rigor and relevance. We removed duplicate records, excluded documents without keywords, and filtered out irrelevant studies, including those labeled with the keyword “current,” which fell outside the defined scope. After applying these refinements, the final dataset consisted of 580 documents suitable for in-depth bibliometric and thematic analysis.

In the data extraction and categorization stage, we developed a classification framework aligned with the research questions in Table 1. We re-crawled the dataset to separate journal and conference publications and applied VOSviewer to conduct co-authorship, co-citation, and keyword co-occurrence analyses. We also used word cloud techniques to identify dominant terms and thematic emphasis. In addition, we employed Latent Dirichlet Allocation topic modeling through the Orange Data Mining platform [12] to uncover latent research themes and examine their evolution over time. This combination of analytical techniques allowed us to systematically address research questions related to topic development, influential contributors, and emerging research directions.

Finally, in the analysis and mapping stage, we performed statistical analyses and constructed visual knowledge maps to synthesize the results. We generated publication trend graphs, keyword networks, thematic clusters, and author impact maps to provide a clear and comprehensive representation of the research landscape. These visualizations facilitated deeper interpretation of structural relationships, research dynamics, and future opportunities in explainable artificial intelligence for financial services. Through this integrated methodological approach, this paper delivered a robust and systematic bibliometric overview of the field.

3. Result and Analysis

According to the review, the search results answer question 1. This trend indicates that between 2018 and 2023, the number of articles published in journals indexed by Scopus will rise significantly. Although the total number may be higher than in other years, there is a decline in 2024 due to publications that still need to make it to the year's end. Researchers are drawn to study XAI, particularly in the financial services industry, and their findings are published as Fig. 1.

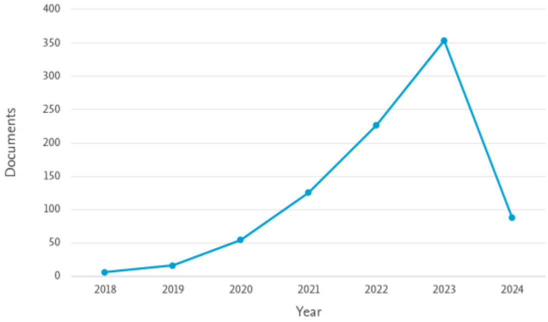


Fig. 1. Number of publications per year

Fig. 2 displays the breakdown of publications based on the type of paper. When compared to other forms of documentation, journal articles and conference proceedings continue to take the lead. At the same time, Fig 3 shows how journal and conference papers are distributed. Publications at conferences have increased dramatically, whereas journal publications have increased slightly. Journal publications may be down in 2023, but that could be because some studies are still in the works and not ready to be published. Publications from high-income nations dominate XAI papers, particularly in the financial services sector (Fig. 4).

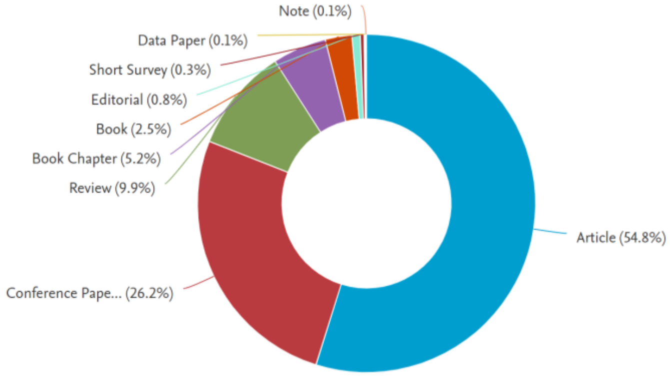


Fig. 2. Publication Distribution by Document Type

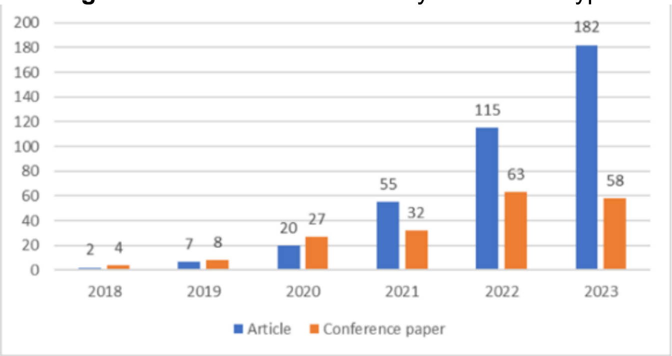


Fig. 3. Annual publication count for journal and conference document types

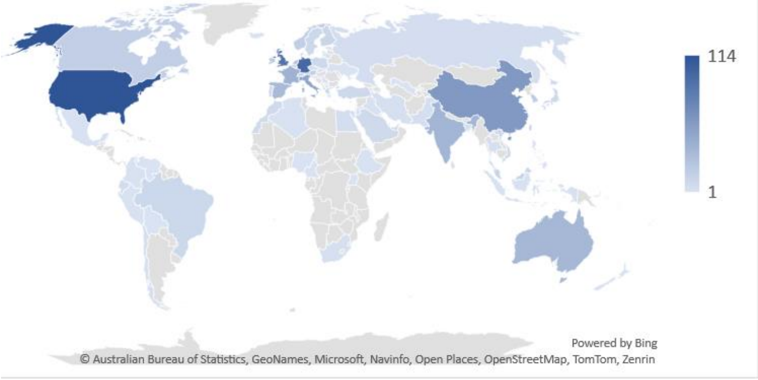


Fig. 4. Publication Distribution by Country

In this study, we address question 2 using VOSviewer version 1.6.19 and WordClouds (https://www.wordclouds.com/). Fig. 5 shows the bibliometric relationships between themes and XAI, including case studies, algorithms, methodologies, and XAI aims. Finance and Credit Scoring is one of the XAI-related case studies. Nodes in the Finance and Credit Scoring domains are colored to show that their respective fields are still in their early stages of research. Fig. 6 displays the keywords that appear at a substantial frequency.

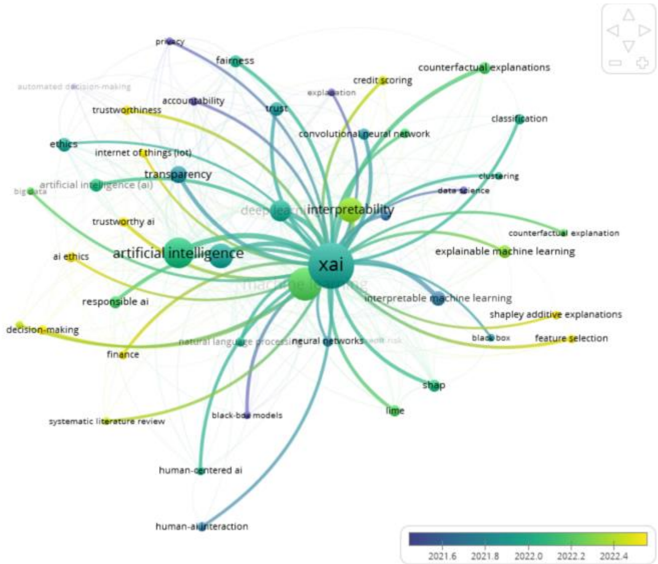


Fig. 5. VosViewer study of bibliographies

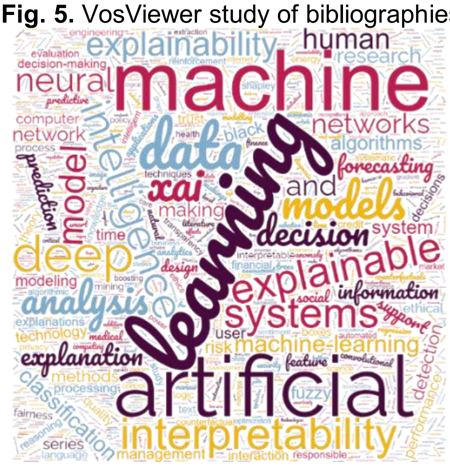
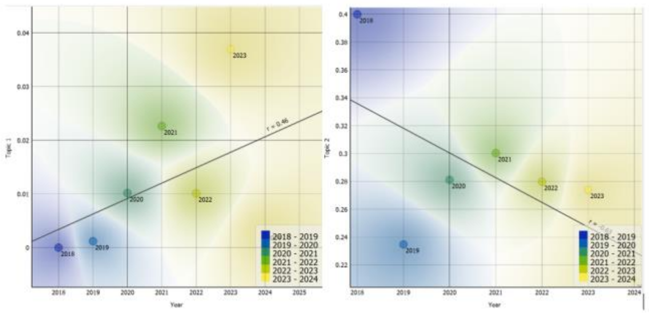
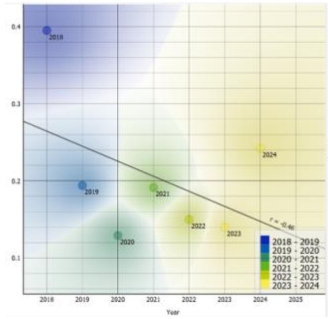


Fig. 6. Analysis of keywords

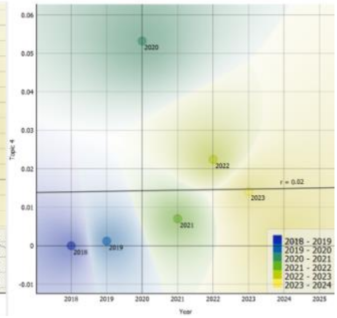
Examining the trend for topics with specific keywords might provide insight into whether a topic is developing. Fig. 7 shows the patterns of each topic keyword generated (Table 2), which answers question 3.



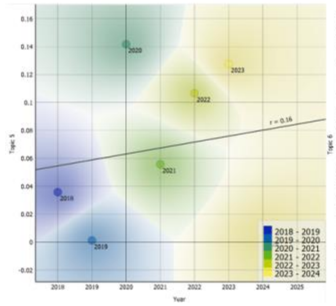
(a) (b)



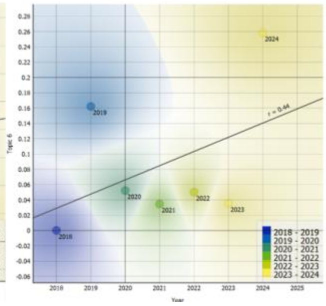
(c)



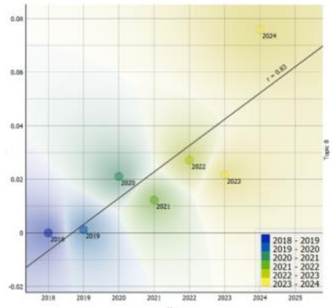
(d)



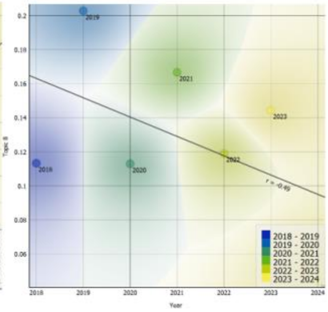
(e)



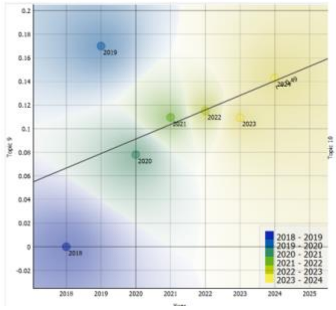
(f)



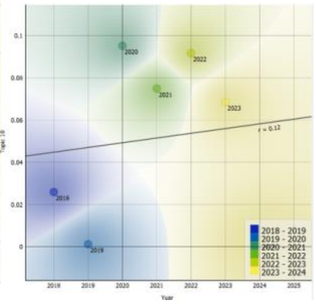
(g)



(h)



(i)



(j)

Fig. 7. The Trend Topic Model

All but one of the themes (b) trend upwards, while topics c and h are trending downwards. The selected scope of XAI for financial services is consistent with topics d, g, and i. This allows us to identify the topic models that answer question 4, as Table 3 shows.

Table 2. Topic Model.

Topic Model	Keyword
a	Energy, learning, risk, machine, decision, performance, systems, model, theory, peer
b	Learning, machine, intelligence, artificial, deep, explainable, systems, neural, decision, model
c	Intelligence, artificial, explainable, AI, systems, learning, user, data, human, analysis
d	Interpretability, financial, trading, knowledge, decision, feature, platform, reasoning, extraction, hoc
e	AI, explainable, learning, machine, intelligence, artificial, systems, networks, neural, model
f	Artificial, intelligence, decision, model, making, study, systems, ethics, transparency, information
g	Learning, self, classification, machine, interpretability, deep, intelligence, banking, artificial, information
h	Decision, AI, making, human, machine, systems, learning, explainable, intelligence, artificial
i	Learning, data, series, time, machine, analysis, model, artificial, explainable, interpretability
j	Learning, detection, machine, data, anomaly, models, representation, explanation, interpretability, deep

Table 3. Promising Topic Keyword.

Topic Model	Keyword
d	Interpretability, financial, trading, knowledge extraction, decision, feature extraction, platform, reasoning
g	information classification, machine learning, interpretability, deep learning, artificial intelligence, banking
i	Data analysis, time series, machine learning, model, artificial, explainable, interpretability

The solutions for the fifth question are in Fig. 8 and Fig. 9. These numbers represent the publishing houses that have published a lot of literature about XAI in the banking sector. Researchers can use this as a starting point for XAI-related searches or publications.

Fig. 8. Publisher Analysis

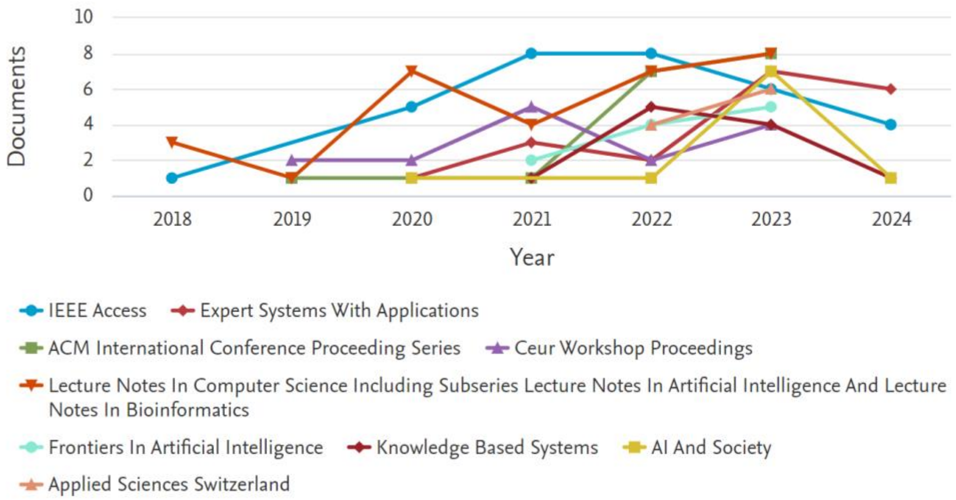


Fig. 9. Ten Publishers with the largest number of publications

As mentioned in question 6, Fig. 10 displays authors who have received several citations. When researching XAI, these writers can be consulted as references.

Fig. 10. Ten authors with the most citations

Our content analysis on recent clusters shows a growing trend in unsupervised learning, especially the application of Autoencoders for fraud detection where labelled data is scarce, even if supervised models (e.g., XGBoost, Random Forest) dominate the landscape (Fig. 11).

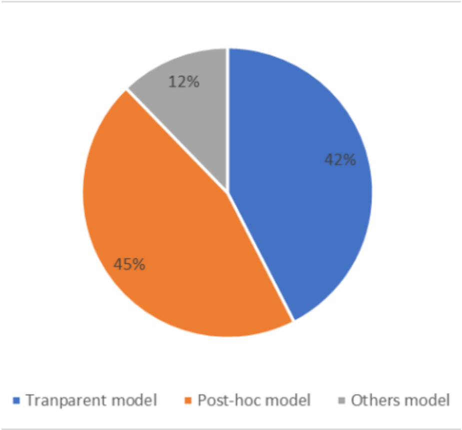


Fig. 11. Variety of machine learning models

Data about researchers' XAI models and methods were also retrieved. Models associated with XAI are categorized in [3] as transparent or post hoc machine learning. Transparent Machine learning models enable users to comprehend the rules used to generate the output. Included in the group of researchers who employ transparent models are [13], [14], [15]. On the other hand, because post hoc models' rules aren't as clear, they are either less clear or need more explanation when talking about the outcomes of specific training procedures. According to references [13], [16], [17], this model is used by researchers. In other cases, however, researchers may offer a model that defies categorization inside machine learning models; for example, see [17], [18], [19], [20], [21], [22], [23]. The abundance of post hoc models, as illustrated in Fig. 11's distribution of model adoption, highlights the critical importance of explainability. Fig. 12 depicts the many strategies applied to the necessity for explainability.

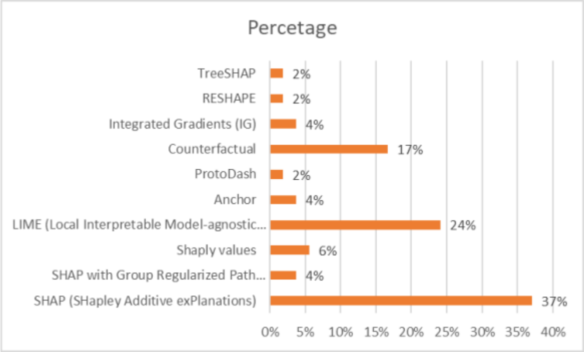


Fig. 12. Distribution of explainable techniques used

A qualitative analysis of the most recent high-impact articles shows a shift towards unsupervised and generative models, despite bibliometric data showing the dominance of post-hoc techniques like SHAP in supervised learning tasks (Fig. 12). In particular, researchers are increasingly using Generative Adversarial Networks (GANs) to handle data imbalance in credit scoring and Autoencoders for anomaly-based fraud detection. These sophisticated applications are compiled and contrasted with conventional review studies in Table 4.

Table 4. Key Studies and Reviews on Advanced XAI Techniques in Finance

Author (Year)	Domain	Method/Focus	Finding/Contribution
Arsenault et al. (2025) [6]	Time Series	Deep Learning Survey	This report provides a detailed analysis of XAI methods developed specifically for complex time series forecasting models.
Khan et al. (2025) [2]	General Finance	Model-Agnostic Methods	Discussing the limitations of current agnostic methods.
Sisodia & Khare (2024)[24]	Market Trend Forecasting	Transformer & Attention	demonstrates how the Transformer model with the Attention mechanism outperforms the traditional LSTM in stock prediction while providing a more understandable feature weight visualisation.
Amirineni S (2025) [25]	Fraud Detection	Unsupervised (Autoencoder)	Proposes a XAI framework that maintains transparency for regulatory compliance while identifying complicated financial fraud using a modern black-box technique.
Gao et al. (Cited in [26], 2023)	Credit Scoring	GAN (Generative AI)	Generative Adversarial Networks (GAN) are used to balance uneven credit data before XAI techniques are used to interpret it.
Kuiper et al. (2022)[4]	Banking Supervision	Stakeholder Perspective	Perspectives from banks vs. supervisory authorities.

4. Conclusion

This study examined the development of explainable artificial intelligence research in financial services through a comprehensive bibliometric and content-based analysis. We found that publication activity increased significantly between 2018 and 2023, indicating a rapidly growing academic interest in XAI, particularly within finance-related applications such as banking, credit scoring, and fraud detection. Although the number of publications appeared to decline in 2024, we interpreted this reduction as a temporal artifact caused by incomplete indexing rather than a true decrease in research activity. We also observed that journal articles and conference proceedings dominated the publication landscape, with conference outputs growing more rapidly, reflecting the fast-paced and evolving nature of

XAI research. Furthermore, publications were largely concentrated in high-income countries, suggesting disparities in research capacity and highlighting opportunities for broader global participation.

By applying VOSviewer, word cloud analysis, and topic modeling techniques, we identified the intellectual structure and thematic evolution of XAI in financial services. We found that core research themes revolved around interpretability, decision-making, machine learning, and financial applications, with specific emphasis on trading, banking, time-series analysis, and data-driven decision support. Most topic models exhibited an upward trend, confirming sustained growth, while only a small subset showed declining interest. Based on keyword co-occurrence and trend analysis, we identified promising research directions that included interpretable financial trading systems, explainable banking classification models, and explainable time-series learning. These findings demonstrated that XAI research is progressively shifting from generic explainability discussions toward domain-specific, application-driven solutions in finance.

In terms of methodological trends, we observed that post hoc explainability techniques remained dominant, particularly in supervised learning models such as Random Forest and XGBoost, underscoring the ongoing need to explain complex black-box systems. However, our qualitative analysis revealed a clear emerging shift toward unsupervised and generative approaches, including Autoencoders and Generative Adversarial Networks, especially for fraud detection and credit scoring under data imbalance and labeling constraints. We also identified influential authors and publishers that shaped the field, providing a valuable reference point for future research. Overall, this paper demonstrated that XAI in financial services has matured into a multifaceted research domain, balancing regulatory demands, methodological innovation, and practical applicability, while opening new avenues for advanced explainable models aligned with real-world financial challenges.

Acknowledgment

This research was supported by Telkom University

References

[1] Grand View Research, “Artificial intelligence market size, share & trends analysis report by solution, by technology (deep learning, machine learning), by end-use, by region, and segment forecasts, 2023–2030,” Grand View Research, 2023. Accessed: Dec. 13, 2023. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market

[2] F. S. Khan, S. S. Mazhar, K. Mazhar, D. A. AlSaleh, and A. Mazhar, “Model-agnostic explainable artificial intelligence methods in finance: A systematic review, recent developments, limitations, challenges, and future directions,” *Artificial Intelligence Review*, vol. 58, no. 8, Aug. 2025, doi: 10.1007/s10462-025-11215-9.

[3] Barredo Arrieta *et al.*, “Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[4] Kuiper, M. van den Berg, J. van der Burgt, and S. Leijnen, “Exploring explainable AI in the financial sector: Perspectives of banks and supervisory authorities,” in *Artificial Intelligence and Machine Learning*. Springer International Publishing, 2022, pp. 105–119, doi: 10.1007/978-3-030-93842-0_6.

[5] D. B. Vuković, S. Dekpo-Adza, and S. Matović, “AI integration in financial services: A systematic review of trends and regulatory challenges,” Springer Nature, Dec. 2025, doi: 10.1057/s41599-025-04850-8.

[6] P. D. Arsenault, S. Wang, and J. M. Patenaude, "A survey of explainable artificial intelligence (XAI) in financial time series forecasting," *ACM Computing Surveys*, vol. 57, no. 10, May 2025, doi: 10.1145/3729531.

[7] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Future Healthcare Journal*, vol. 6, no. 2, pp. 94–98, Jun. 2019, doi: 10.7861/futurehosp.6-2-94.

[8] R. M. Vasse'i and J. McCrosky, "AI transparency in practice: Builders on what works—and what doesn't," Mozilla Foundation, 2023. Accessed: Dec. 13, 2023. [Online]. Available: https://foundation.mozilla.org/en/research/library/ai-transparency-in-practice/ai-transparency-in-practice/

[9] J. Černevičienė and A. Kabašinskas, "Review of multi-criteria decision-making methods in finance using explainable artificial intelligence," *Frontiers in Artificial Intelligence*, vol. 5, Mar. 2022, doi: 10.3389/frai.2022.827584.

[10] M. Salama, R. Bahsoon, and N. Bencomo, "Managing trade-offs in self-adaptive software architectures: A systematic mapping study," in *Managing Trade-Offs in Adaptable Software Architectures*, I. Mistrik *et al.*, Eds. Boston, MA, USA: Morgan Kaufmann, 2017, pp. 249–297, doi: 10.1016/B978-0-12-802855-1.00011-3.

[11] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and Software Technology*, vol. 64, pp. 1–18, 2015, doi: 10.1016/j.infsof.2015.03.007.

[12] University of Ljubljana, "Topic modelling," Orange Data Mining. Accessed: Mar. 23, 2024. [Online]. Available: https://orangedatamining.com/widget-catalog/text-mining/topicmodelling-widget/

[13] J. Dessain, N. Bentaleb, and F. Vinas, "Cost of explainability in AI: An example with credit scoring models," in *Communications in Computer and Information Science*, vol. 1901, 2023, doi: 10.1007/978-3-031-44064-9_26.

[14] M. Pamuk and M. Schumann, "Opening a new era with machine learning in financial services? Forecasting corporate credit ratings based on annual financial statements," *International Journal of Financial Studies*, vol. 11, no. 3, Sep. 2023, doi: 10.3390/ijfs11030096.

[15] L. M. Demajo, V. Vella, and A. Dingli, "Explainable AI for interpretable credit scoring," in *Computer Science & Information Technology (CSIT)*, AIRCC, Nov. 2020, pp. 185–203, doi: 10.5121/csit.2020.101516.

[16] R. Chimatapu, H. Hagras, M. Kern, and G. Owusu, "Hybrid deep learning type-2 fuzzy logic systems for explainable AI," in *Proc. IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE)*, 2020, pp. 1–6, doi: 10.1109/FUZZ48607.2020.9177817.

[17] L. O. Hjelkrem and P. E. de Lange, "Explaining deep learning models for credit scoring with SHAP: A case study using open banking data," *Journal of Risk and Financial Management*, vol. 16, no. 4, Apr. 2023, doi: 10.3390/jrfm16040221.

[18] V. La Gatta, V. Moscato, M. Postiglione, and G. Sperlì, "CASTLE: Cluster-aided space transformation for local explanations," *Expert Systems with Applications*, vol. 179, Oct. 2021, doi: 10.1016/j.eswa.2021.115045.

[19] W. Li, F. Paraschiv, and G. Sermpinis, "A data-driven explainable case-based reasoning approach for financial risk detection," *Quantitative Finance*, vol. 22, no. 12, pp. 2257–2274, 2022, doi: 10.1080/14697688.2022.2118071.

[20] Š. Lyócsa, P. Vašaničová, B. H. Misheva, and M. D. Vateha, "Default or profit scoring credit systems? Evidence from European and U.S. peer-to-peer lending markets," *Financial Innovation*, vol. 8, no. 1, Dec. 2022, doi: 10.1186/s40854-022-00338-5.

[21] M. Argotty-Erazo *et al.*, "A novel linear-model-based methodology for predicting the directional movement of the euro–dollar exchange rate," *IEEE Access*, vol. 11, pp. 67249–67284, 2023, doi: 10.1109/ACCESS.2023.3285082.

[22] Ghosh, E. Alfaro-Cortés, M. Gámez, and N. García-Rubio, "Role of COVID-19 media chatter in predicting the Indian stock market: An integrated framework of nonlinear feature transformation and advanced AI," *Expert Systems with Applications*, vol. 219, p. 119695, 2023, doi: 10.1016/j.eswa.2023.119695.

[23] T. Buğra, Ö. İcan, and E. Bulut, "Extending machine learning prediction capabilities by explainable AI in financial time series prediction," SSRN, Jul. 2022, doi: 10.2139/ssrn.4170455.

[24] L. S. Sisodia and A. Khare, "Enhancing market trend forecasting with explainable AI: A comparative analysis of deep learning models and interpretability techniques," *ShodhKosh: Journal of Visual and Performing Arts*, vol. 5, no. 3, pp. 1712–1722, 2024, doi: 10.29121/shodhkosh.v5.i3.2024.61.

[25] S. Amirineni, “Explainable artificial intelligence (XAI) models for transparent and accountable fraud detection in banking ecosystems,” *International Journal of Scientific Research and Management*, vol. 13, no. 8, pp. 2493–2513, Aug. 2025, doi: 10.18535/ijsrm/v13i08.ec03.
[26] F. M. Talaat *et al.*, “Toward interpretable credit scoring: Integrating explainable artificial intelligence with deep learning for credit card default prediction,” *Neural Computing and Applications*, 2023, doi: 10.1007/s00521-023-09232-2.
[27] M. Goram, “Towards a standardized interactive AI transparency label,” in *Intelligent Systems and Applications*. Springer Nature Switzerland, 2025, pp. 539–547.
[28] G. Chakkappan, A. Morshed, and M. M. Rashid, “Explainable AI and big data analytics for data security risk and privacy issues in the financial industry,” in *Proc. IEEE Conf. on Engineering Informatics (ICEI)*, 2024, pp. 1–9, doi: 10.1109/ICEI64305.2024.10912422.
[29] H. Baradaran, M. Bohlouli, and M. R. J. Motlagh, “Decoding tomorrow’s gold prices: A comparative study of GRU and CNN–LSTM models in the Iranian market,” in *Proc. 10th Int. Conf. on Web Research (ICWR)*, IEEE, 2024, pp. 329–334.
[30] W. K. M. Woon *et al.*, “Prediction of gold prices using a hybrid ARIMA–LSTM model,” *Advances in Applied Statistics*, vol. 92, no. 5, pp. 749–766, 2025.
[31] Kashif and R. Ślepaczuk, “LSTM–ARIMA as a hybrid approach in algorithmic investment strategies,” *Knowledge-Based Systems*, vol. 323, p. 113563, 2025, doi: 10.1016/j.knosys.2025.113563.
[32] Mohamed, K. Abdelqader, and K. Shaalan, “Explainable artificial intelligence: A systematic review of progress and challenges,” *ISWA Journal*, 2025, doi: 10.1016/j.iswa.2025.200595.
[33] F. Almalki and M. Masud, “Financial fraud detection using explainable AI and stacking ensemble methods,” *arXiv preprint* arXiv:2505.10050, 2025.
[34] R. Ye and J. Chen, “Unlocking the black box: A five-dimensional framework for evaluating explainable AI in credit risk,” *arXiv preprint* arXiv:2511.04980, 2025.