International Journal of Information Management Data Insights 4 (2024) 100234

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

How can artificial intelligence help customer intelligence for credit portfolio management? A systematic literature review

Alessandra Amato [a], Joerg R. Osterrieder [b,c], Marcos R. Machado [b,*]

[a] University of Twente, Faculty of Electrical Engineering, Mathematics, and Computer Science, AE Enschede, 7500, Netherlands
[b] University of Twente, Faculty of Behavioural, Management and Social Sciences, Department of High-Tech Business and Entrepreneurship, AE Enschede, 7500, Netherlands
[c] Bern Business School, Institute of Finance and Applied Data Science, Bern, 3005, Switzerland

ARTICLE INFO

Keywords:
Early warning systems
Customer segmentation
Lending settings
Unsupervised learning
Systematic literature review

ABSTRACT

In this era of Big Data and the advancement of sophisticated analytical techniques, the financial industry has the capacity to implement innovative technologies within their systems to derive crucial insights about their clientele and vigilantly monitor their activities. This landscape has seen the emergence and rise of two significant applications, namely, customer segmentation systems and early warning systems. Therefore, this study presents a systematic literature review on the automation of customer segmentation and early warning techniques with a focus on managing credit portfolio entities. The research delves into a multitude of scholarly articles from three distinct perspectives: charting the dominant trends within the literature, unpacking the overarching themes, and critically examining the integration of early warning signals within customer clustering applications. Furthermore, the review reveals a noticeable dearth of studies probing the synergistic application of these two systems. Despite their independent effectiveness in risk management and targeted marketing strategies respectively, an integrated approach holds potential for bolstering financial stability and tailoring customer service. Thus, this review stands as a significant academic contribution, advocating an integrated application of these systems within the financial industry. The findings provide a novel foundation for future research and practical applications, potentially redefining strategies within the financial sector.

1. Introduction

This study draws its foundations from the unique interplay of Customer Segmentation and Early Warning Systems at a large international banking corporation. The exploration of this dynamic relationship aims to unveil new paths for enhancing business profitability, primarily by leveraging early warning triggers to identify distinctive customer clusters, opening avenues for more nuanced and targeted marketing initiatives and commercial endeavors.

In today's competitive lending market, maintaining a competitive edge and enhancing profitability are the main objectives for financial institutions. With the advent of Big Data and the surge of advanced analytic techniques, these institutions are now capable of deriving meaningful insights from large, diverse datasets, enabling the development of more targeted strategies (Ram et al., 2016). Among the array of data-driven methods, Customer Segmentation (CS) has become a highly effective approach. CS involves dividing a broad customer base into smaller subsets, with each sharing similar characteristics relevant to an organization's marketing and sales objectives (ECB, 2017).

In the context of the financial sector, specifically in the lending field, CS has proven instrumental in offering tailored products and services based on individual customer segments, thereby improving risk management procedures (Raiter, 2021). Furthermore, the deployment of CS allows marketers to readily identify cross and up-selling opportunities. However, the success of segmentation is significantly influenced by the selection of features for building clusters (Mousaeirad, 2020). This critical step necessitates comprehensive business understanding and consideration of the initial business objectives leading to the application development (Mihova & Pavlov, 2018). Practical examples of how the loan industry can leverage CS are evident when banks and financial institutions undertake risk assessments, set interest rates, customize loan products, and formulate marketing strategies (Yuping et al., 2020, Bach et al., 2013, Zand, 2020, Herrmann & Masawi, 2022). By categorizing consumers according to credit history, income, employment status, and

* Corresponding author.
E-mail addresses: a.amato@student.utwente.nl (A. Amato), joerg.osterrieder@utwente.nl (J.R. Osterrieder), m.r.machado@utwente.nl (M.R. Machado).

https://doi.org/10.1016/j.jjimei.2024.100234

Available online 15 April 2024

other financial determinants, lenders are equipped to more accurately evaluate the risk associated with lending. This segmentation enables entities to propose lower interest rates to segments perceived as lower risk, while imposing higher rates on those deemed higher risk. Furthermore, financial organizations employ customer segmentation to devise loan offerings that cater to the specific requirements of distinct customer groups. For instance, young professionals might be targeted with products like first-time homebuyer loans or loans aimed at career development, while retirees could be presented with reverse mortgages or equity release schemes. Marketing strategies are then tailored to each segment, employing particular language, mediums, and messages that align with the preferences and behaviors of each group. This tailored approach not only heightens the efficacy of marketing initiatives but also bolsters customer engagement and loyalty by delivering more pertinent and attractive propositions. It is also important to highlight the slight difference between CS and clustering (Kashwan & Velu, 2013, Kansal et al., 2018, Herrmann & Masawi, 2022). The former is a more deliberate process, where entities categorize clients into predefined groups based on specific criteria, while the latter involves grouping clients based on similarities in financial health and risk-related indicators without prior knowledge of the groups. In the finance industry, especially for credit portfolio entities seeking potential business-making prospects via client clusters, it is vital to incorporate both financial health and risk-related indicators to obtain a nuanced and reliable client base representation.

Parallel to CS, Early Warning Systems (EWS), qualitative and quantitative indicators forecasting risk events, have seen significant growth in the past decade, becoming an invaluable asset for credit risk monitoring. Despite remaining an untapped field for many enterprises, an increasing number of organizations have begun recognizing their potency, integrating EWS into their risk management practices, especially in the financial sector (Klopotan et al., 2018).

The integration of information from EWS within CS systems may offer a valuable enhancement to these processes. The inclusion of the risk factor when dealing with insolvent clients would provide insights on the current and prospective credit health of portfolio entities for each segment. Accordingly, this study presents a systematic literature review on the automated CS techniques and EWS for credit portfolio monitoring in financial lending settings, also exploring the integration of early warning signals within customer clustering processes to derive even more insightful customer segments.

By gathering relevant evidence on these two primary topics that fulfill pre-defined eligibility criteria, this study seeks to answer the following research questions:

1. What are the Early Warning Signals and Customer Segmentation frameworks used in credit portfolio monitoring?
2. How can EWS and CS methods be validated, and how are these methods associated with each other?

The insights gleaned from this systematic review offer multifaceted benefits to both corporate entities and scholarly researchers. On one hand, financial industry professionals can leverage these findings to deepen their understanding of the risk-mitigating and strategic potential of EWS and CS systems. This enhanced understanding paves the way for devising more targeted and efficient marketing strategies, fostering customer retention, and practically implementing these two pioneering technologies.

On the other hand, this study will significantly contribute to the advancement of research in these two somewhat nascent fields. By elucidating the underlying mechanisms driving CS and EWS, the study can spark the development of novel theories and methodologies. This will enable organizations to maintain their competitive edge in an evolving market landscape by staying at the forefront of innovation.

The paper is structured as follows: Section 2 outlines the methodology deployed to carry out the systematic literature review, detailing the steps undertaken to extract and scrutinize the relevant articles. Section 3, delves into the principal trends identified within the literature for both CS and EWS models. Subsequently, Section 5 offers a detailed overview of the emerging and recurring themes identified within the reviewed articles. Section 6 discusses the potential interconnections between the two concepts and their potential for integration. Finally, Section 7 provides a comprehensive exploration of the limitations and future recommendations based on the findings of this review, in addition to summarizing the entire review and presenting final thoughts and conclusions.

2. Methodology

In our systematic literature review study, we adhere to a structured process to ensure comprehensive and unbiased outcomes. Drawing upon the framework outlined by S et al. (2024), our methodology encompasses eight essential steps, which sets the direction and scope of our review. Initially, we establish the research problem (research questions). Following this, we create and validate a review methodology to provide guidance for our data search and analysis. The third step entails a comprehensive analysis and exploration of the pertinent literature, guaranteeing a comprehensive exploration of the subject matter. Subsequently, we employ specific criteria to selectively screen the gathered studies for inclusion, so guaranteeing their pertinence and excellence. The fifth phase entails a meticulous assessment of the quality of each chosen study. Subsequently, the process of data extraction is initiated, wherein relevant information is methodically collected from each study. Data analysis and synthesis, as the seventh phase, enables us to integrate findings and derive significant insights. Ultimately, we condense the outcomes into comprehensive reports, showcasing our findings in a methodical and easily understandable fashion. The employed methodological approach guarantees the rigor and reliability of our literature review, thereby providing important insights into the topic matter.

Scopus[1] was selected as the primary database for this systematic literature review, owing to its extensive repository of scientific articles and sophisticated search features (Kushwaha et al., 2021, Ngai et al., 2009). The advanced search function was utilized to create custom search queries, featuring keywords including: ‘Early Warning Systems’, ‘Credit Risk’, ‘Financial Distress’, ‘Customer Segmentation’, ‘Customers Clustering’, ‘Unsupervised Machine Learning’, ‘Credit Portfolio Monitoring’, ‘Lending’, and ‘Loans’. These keywords, in combination with logical operators ‘AND’ and ‘OR’, facilitated the crafting of three search queries, given the research’s focus on the application of EWS in credit risk monitoring and CS in credit portfolio management:

- “Early Warning Systems” AND (“Credit Risk” OR “Financial Distress” OR “Lending” OR “Bankruptcy”)
- (“Customer Segmentation” OR “Customers Clustering” OR “Clustering” OR “Unsupervised Machine Learning” OR “Unsupervised Learning”) AND (“Credit Portfolio Monitoring” OR “Lending” OR “Loans” OR “Loan”)
- “Early Warning Systems” AND “Customer Segmentation”

These queries targeted keywords, abstracts, and titles of articles in the initial retrieval phase. These queries were applied to Scopus, and the resulting papers were filtered based on the exclusion criteria outlined in Table 1. Using a four-phase strategy, we gathered 49 publications through the methodical method outlined in this chapter and depicted in Fig. 1. Furthermore, publications from reputable journals in the field that were not retrieved by the search query were manually included to offer a comprehensive and benchmark perspective on the methodologies discussed in this literature review. The total number of publications analyzed in this study is 67.

[1] https://www.scopus.com/home.url.

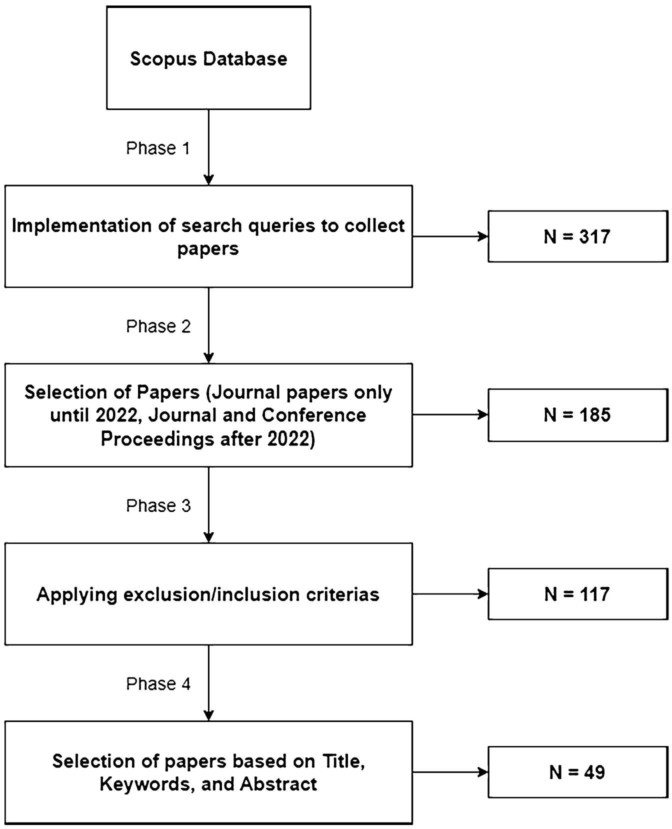

Fig. 1. Illustration of the article selection process.

Table 1
Summary of article selection criteria.

Criteria	Decision
Inclusion of pre-defined keywords in title, abstract, or keyword list	Inclusion
Article publication in a scientific journal	Inclusion
Article written in English	Inclusion
Article published before 2017	Exclusion
Duplicates of an original article	Exclusion
Relevance of abstract, title, and content to research objective	Exclusion
Unavailability of the article online for free	Exclusion

The rigorous methodology employed has ensured a robust and highly relevant collection of articles for this systematic review.

3. Deconstructing the research landscape

This section is structured into three parts, each providing a unique lens to examine the scholarly work surrounding the application of CS techniques in lending and the use of EWS for credit risk monitoring.

The first subsection delves into the temporal distribution of the related literature, offering a timeline of the scholarly contributions to these fields. This analysis will help to identify potential trends, patterns or shifts over the years, giving a historical context to the current state of research.

In the second subsection, we turn our attention to the sources of this literature by assessing the distribution across different journals. Understanding this distribution not only showcases the breadth of the research landscape but also highlights the most active and influential platforms fostering discourse on these topics.

Finally, in the third subsection, we investigate the most recurrent keywords within the examined literature. This analysis serves as a thematic compass, pointing to the core themes and concepts that anchor research in these fields, as well as hinting at emerging areas of interest.

To facilitate comprehension and offer a complete perspective, all the literature consulted for this study is indexed in Table A1. This table provides a comprehensive overview of each paper, highlighting critical aspects such as the settings in which the respective model was employed, the main objectives of the research, the data-driven techniques applied, and the evaluation methods utilized to gauge the efficiency of the approach. This detailed catalog not only contributes to the transparency of the research process but also allows readers to trace the intellectual journey that led to the findings of this study.

3.1. Temporal distribution of literature

We begin by examining the temporal progression of the scholarly discourse on CS in lending and EWS for credit risk monitoring. Fig. 2 demonstrates a rise in the number of papers published in recent years, particularly focusing on studies in the field of CS. Although the quantity of studies on EWS has decreased in 2023, the continued relevance of the topic is evident, particularly when considering the absence of any publications on this subject in 2028 and 2020. Therefore, the increase in publications in 2022 and 2023 for CS suggests a possible rise in interest and use of clustering models in financial institutions. An observed shift should also be highlighted in the literature within the finance domain indicates a significant increase in the recognition of EWS's potential.

These temporal trends in the literature not only provide a historical view of the research interests but also hint at the areas gaining momentum, providing valuable insights for future research directions.

3.2. Journal distribution of literature

The next analysis dives into the journal distribution of the collected literature. A comprehensive understanding of this dispersion provides insights into the diversity and reach of the research community addressing these topics.

Table 2 enumerates all CS-related articles, indicating the journal of publication, the number of citations collected, and the impact factor of each journal for 2021, reflecting its most recent score. Two studies (Moradi & Mokhatab Rafiei, 2019, Yuan et al., 2022) stand out in the citation count, suggesting their significant influence in the field. Interestingly, each paper appears to have been published in a unique journal, underscoring the lack of a single preferred journal among researchers in this area.

Following the same pattern, Table 3 highlights the EWS-related articles. A similar diversification across different journals is apparent. Notably, "Mobile Information Systems" has surfaced as a preferred source, contributing the largest share of articles. The citation counts and impact factor data show a trend consistent with the CS-related literature.

The disparate distribution of literature across various journals, along with varying citation counts and impact factors, reflects the multidisciplinary interest and diversified reach of these topics. These trends hint at the fields' widespread relevance and growing influence in shaping the future of lending practices.

3.3. Distribution based on keywords

Here we highlight the analysis of the most frequently occurring keywords within the examined articles, underlining their essential role in simplifying the research process for academics and search engines alike. By facilitating the swift identification of pertinent studies, keywords essentially outline the scope, topics, and core focus of the articles. However, a notable observation is that a significant fraction of the academic papers, particularly those related to EWS, lacked keywords, leading to their exclusion from this specific analysis. Thus, the focus was shifted towards papers that did incorporate keywords, offering insights into prevalent themes and areas of interest.

The investigation into CS-related papers unveiled "credit scoring" as the predominant keyword, followed by "k-means," "clustering," and

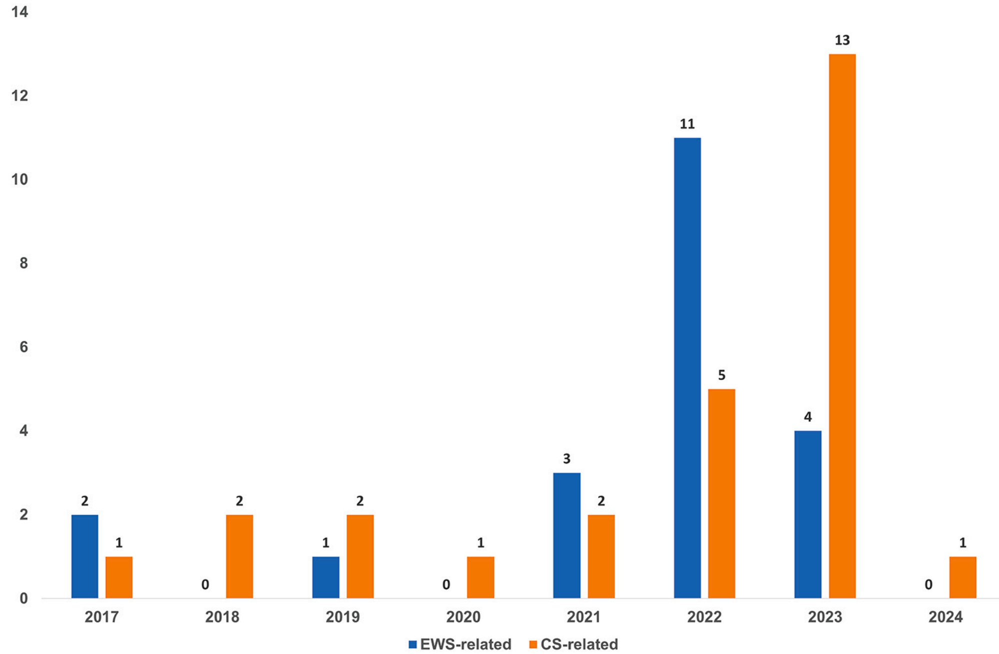

Fig. 2. Temporal distribution of research papers on EWS and CS-related topics.

Table 2
Table reporting the journal distribution of CS-related articles.

Topic	Study	Journal	# Citations	Impact Factor
CS	Kaminskyi et al. (2022)	Journal of Risk and Financial Management	–	2.300
CS	Jadwal et al. (2022)	Microsystem Technologies	3	2.012
CS	Machado and Karray (2022a)	Electronic Commerce Research and Applications	–	5.622
CS	Tasgetiren et al. (2022)	Concurrency and Computation Practice and Experience	1	1.831
CS	Yuan et al. (2022)	Research in International Business and Finance	10	6.143
CS	Pandey and Shukla (2021)	Reliability: Theory and Applications	–	0.440 –
CS	Singh et al. (2021)	Applied Soft Computing	5	8.263
CS	Lazo et al. (2020)	Journal of Credit Risk	1	0.880
CS	Moradi and Mokhatab Rafiei (2019)	Financial Innovation	42	–
CS	Nazari et al. (2019)	International Journal of Supply Chain Management	–	–
CS	Philip et al. (2018)	Data Base for Advances in Information Systems	5	1.828
CS	Firouzabadi et al. (2018)	International Journal of Electronic Customer Relationship Management	–	–
CS	Luthfi and Wibowo (2017)	International Journal of Simulation: Systems, Science and Technology	–	–
CS	Kita et al. (2023)	Cultural Management: Science and Education	1	0.2300
CS	Parmar et al. (2023)	2023 4th International Conference on Electronics and Sustainable Communication Systems	–	–
CS	Lenka et al. (2023)	Risk Management	–	2.560
CS	Solimun et al. (2023)	Advanced Mathematical Models and Applications	–	0.340
CS	Liu et al. (2023)	Tehnicki Vjesnik	–	0.900
CS	Huang et al. (2023)	Annals of Operations Research	20	4.800
CS	Boyapati and Aygun (2023)	17th IEEE International Conference on Semantic Computing	–	–
CS	Gorle and Panigrahi (2023)	Multimedia Tools and Applications	–	2.577
CS	Vlahavas et al. (2024)	Financial Innovation	–	8.400
CS	Chen et al. (2023)	Journal of Management Science and Engineering	5	6.600
CS	Han et al. (2023)	Knowledge and Information Systems	–	2.700
CS	Ahelegbey and Giudici (2023)	Risks	–	2.200
CS	Yin et al. (2023)	Applied Soft Computing	4	8.263
CS	Mayorova et al. (2023)	Financial and Credit Activity: Problems of Theory and Practice	–	–

“anfis,” indicating these topics as recurring themes within the literature. These keywords suggest a concentrated interest in specific methodologies and applications within the credit scoring domain. On the contrary, the analysis of EWS-related articles revealed a different set of dominant keywords, namely “machine learning,” “early warning systems,” “financial crisis,” and “deep learning.” This contrast highlights the varied focal points between the two fields, with EWS-related research emphasizing technological and systemic approaches to predicting financial crises.

Figs. 3 and 4 present a word cloud showcasing the most prevalent keywords in the studied CS- and EWS-related papers. The word size in the cloud is indicative of the relative frequency of each keyword in the literature. As we can notice, the description earlier provided shows that, for CS, “credit scoring” emerges as the most commonly used keyword, trailed by “k-means”, “clustering”, and “anfis”. This finding implies that these subjects were discussed in multiple articles, hence they could be recurrent themes in this sphere. Regarding EWS, the most prominent keywords that recur multiple times are “machine learning”, “early warning systems”, “financial crisis”, and “deep learning”.

4. The role of AI in credit portfolio management

The emergence of AI has brought substantial changes in the field of credit portfolio management within the financial services indus-

Table 3
Table reporting the journal distribution of EWS-related articles.

Topic	Study	Journal	# Citations	Impact Factor
EWS	Wang and Zhang (2023)	Information Processing Management	–	7.466
EWS	Guerra et al. (2022)	Economic Analysis and Policy	4	4.444
EWS	Petropoulos et al. (2022)	Intelligent Systems in Accounting, Finance and Management	2	–
EWS	Wangsong (2022)	Mobile Information Systems	–	–
EWS	Xie and Shi (2022)	Mobile Information Systems	–	–
EWS	Han (2022)	Computational Intelligence and Neuroscience	–	–
EWS	Yin et al. (2022)	Computational Intelligence and Neuroscience	2	–
EWS	Xie (2022)	Mobile Information Systems	–	–
EWS	Huang et al. (2022)	Annals of Operations Research	4	4.820
EWS	Xu et al. (2022)	Annals of Operations Research	5	4.820
EWS	Tong and Tong (2022)	Scientific Programming	1	–
EWS	Yang (2022)	Computer-Aided Design and Applications	1	–
EWS	Jacobs (2021)	International Journal of Financial Studies	2	–
EWS	Zhu et al. (2021)	Information Processing and Management	21	7.466
EWS	Aytaç Emin et al. (2021)	Applied Economics Letters	1	1.287
EWS	Zhang et al. (2019)	IEEE Access	2	3.476
EWS	Pompella and Dicanio (2017)	Economic Modelling	7	3.875
EWS	Berlinger (2017)	Finance Research Letters	–	9.846
EWS	Hacibedel and Qu (2023)	Journal of Credit Risk	2	0.220
EWS	Ramesh and Jeyakarthic (2023)	Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications	–	3.174
EWS	Clintworth et al. (2023)	Maritime Economics and Logistics	15	3.119
EWS	Madhaveelatha et al. (2023)	Lecture Notes in Networks and Systems	–	0.540

Fig. 3. Word cloud of CS-related articles' keywords.

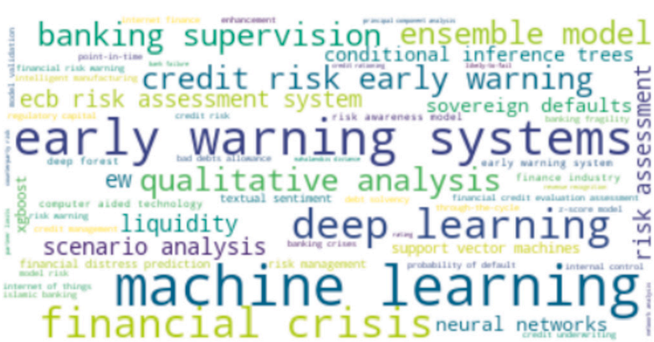

Fig. 4. Word cloud of EWS-related articles' keywords.

try (Machado & Karray, 2022c, Liu et al., 2023, Huang et al., 2023). This includes a variety of important tasks, including assessing credit risk (Machado & Karray, 2022b, Nazari et al., 2019) and segmenting customers (Ahelegbey & Giudici, 2023, Yin et al., 2023, Vlahavas et al., 2024), which are necessary for maximizing profits and minimizing losses. AI technologies, which utilize extensive datasets and advanced algorithms, are leading the way in enhancing these processes. They offer sophisticated methods to explore borrowers' behavior, improve the precision of predictions, and enable tailored financial services, thus transforming conventional methods. The versatility and personalized methods facilitated by AI applications are evident in their various settings, which span from retail banking to corporate finance, with their different market groups and risk profiles.

AI plays a prominent role in credit risk management (Fujii et al., 2022, Pandey et al., 2021, Sasaki et al., 2024). The utilization of linear regressions and traditional statistical methods in conventional models may not comprehensively encompass the intricacies inherent in financial data or the nuanced dynamics of economic cycles. AI, on the other hand, presents a more dynamic and advanced analytical framework by utilizing ML models. Various ML algorithms, including decision trees, support vector machines, and neural networks, have been proved to be capable of revealing intricate patterns and connections within datasets that may not be readily apparent by conventional analytical methods. These models are applied in diverse contexts, ranging from the evaluation of individual borrower risk to the examination of credit risk across entire portfolios.

The utilization of AI for consumer segmentation represents a notable progression, facilitating financial institutions in providing services that are more tailored and efficient (Hu et al., 2023, Talaat et al., 2023). AI models have the capability to handle and examine extensive datasets in order to identify separate client segments using different criteria, including spending patterns, credit usage, and even behavioral aspects. The clustering of customers into distinct segments enables the creation of customized marketing tactics and product offerings, hence enhancing consumer engagement and happiness. Cluster analysis and principle component analysis are often used methods for segmentation, enabling organizations to accurately identify and predict client demands and preferences.

The assessment techniques employed for AI applications in credit portfolio management guarantee the efficacy and dependability of these technologies. Standard performance criteria used to evaluate the predictive capability of AI models in credit risk management include accuracy, precision, recall, and the area under the receiver operating characteristic curve. Likewise, the efficacy of customer segmentation endeavors is frequently assessed by evaluating metrics such as customer engagement, rates of customer retention, and the profitability of focused marketing initiatives (Machado & Karray, 2022b, Thantharate, 2023, Han et al., 2023, Yin et al., 2023, Ramesh & Jeyakarthic, 2023). In addition, it is imperative to verify, validate, and ensure compliance of these models with regulations and user requirements in order to adapt to evolving market dynamics and consumer preferences.

5. Predominant themes in literature

In the pursuit of understanding the prevailing trends and practices within the domains of Early Warning (EWS) and Customer Segmentation (CS) systems, this section examines the significant themes in the contemporary literature. The EWS and CS systems, playing pivotal roles in risk management and customer relationship domains respectively, utilize sophisticated techniques and methodologies to bolster operational efficiency and improve decision-making. Given the ever-increasing complexity and technological progression in these areas, it is crucial to critically examine the predominant themes from various angles to provide comprehensive insight into their operations and efficacy.

To facilitate this, the section is subdivided into three critical subsections, each concentrating on a specific theme concerning EWS and CS systems.

- **Settings Analysis:** This subsection dives into the distinct environments and scenarios in which EWS and CS systems are utilized. By examining the unique settings of the investigated systems, we aim to identify the multitude of parameters and variables that influence the overall system performance and accuracy.
- **Techniques Analysis:** This subsection dissects the assortment of techniques employed in constructing EWS and CS systems. By meticulously scrutinizing different methodologies and algorithms, we aim to understand the diverse technical facets that contribute to these systems' overall functionality and efficacy.
- **Evaluation Methods Analysis:** The final subsection shifts its focus towards the evaluation methodologies prevalent in the literature for assessing EWS and CS system performance. These methodologies form the backbone of model validation and performance benchmarking, underscoring their importance in this context.

5.1. Settings analysis

This initial subsection strives to present a comprehensive snapshot of the frequent settings appearing in the literature associated with both EWS and CS models. The term "settings" encapsulates the contexts in which the studies were conducted, which inevitably shaped the project's requirements and influenced the experiment subjects.

A significant portion of the CS literature was found to be centered on Peer-to-Peer (P2P) lending services, a form of fintech that allows individuals to lend or borrow from others, circumventing traditional financial institutions (Kaminskyi et al., 2022, Jadwal et al., 2022, Machado & Karray, 2022a, Pandey & Shukla, 2021, Ahelegbey & Giudici, 2023, Yin et al., 2023). The Lending Club dataset, which contains demographic and behavioral data from users of the Lending Club digital marketplace, was extensively utilized in various studies to conduct and validate experiments (Jadwal et al., 2022, Pandey & Shukla, 2021). However, it is crucial to highlight that a majority of the articles examined primarily sought to segment clients of banking institutions.

One noteworthy trend unveiled during the settings analysis concerns the type of entities investigated. Retail banking customers, comprised of individuals from the general public, were predominantly the subjects of the clustering models (Pandey & Shukla, 2021, Moradi & Mokhatab Rafiei, 2019, Nazari et al., 2019, Philip et al., 2018, Firouzabadi et al., 2018, Luthfi & Wibowo, 2017, Mayorova et al., 2023, Vlahavas et al., 2024, Chen et al., 2023). The attributes used to generate these clusters often echoed the socio-economic profile of the subjects, making them suitable only for segmenting real physical borrowers.

Lastly, the analysis also underscored that a vast majority of CS articles were deployed for risk assessment and monitoring. Most papers reviewed employed customer clustering techniques to construct models that identify high-risk borrowers with a heightened probability of financial distress. Only a handful of studies used these techniques to attain different objectives, such as enhancing client-product allocation (Firouzabadi et al., 2018).

Shifting focus to the settings analysis in literature pertaining to EWS usage for credit risk control, one significant observation merits discussion. Contrary to the CS techniques, studies related to EWS targeted a diverse array of industries and institutions. While Internet Finance and banking fields were still well-represented (Guerra et al., 2022, Wangsong, 2022, Xie & Shi, 2022, Aytaç Emin et al., 2021, Pompella & Dicanio, 2017, Berlinger, 2017), numerous early warning applications were developed to quantify and evaluate the financial risk of various types of entities, including IoT companies, manufacturing enterprises, and governments (Wang & Zhang, 2023, Petropoulos et al., 2022, Han, 2022, Yin et al., 2022, Xie, 2022, Xu et al., 2022, Zhu et al., 2021, Hacibedel & Qu, 2023, Ramesh & Jeyakarthic, 2023, Clintworth et al., 2023).

As outlined in Table A1, a considerable number of studies were centered on the Chinese market, both from a banking perspective and in the context of broader industries and firms (Wang & Zhang, 2023, Wangsong, 2022, Xie & Shi, 2022, Huang et al., 2022).

5.2. Techniques analysis

The evolution of technology, coupled with the advent of cutting-edge artificial intelligence applications, has paved the way for numerous solutions in crafting robust EWS and CS systems. This subsection sheds light on the primary techniques that have been employed across different studies, with the aim of categorizing the most prevalent methods and enhancing our comprehension of these systems from a technical viewpoint.

Concerning the literature that revolves around CS applications, several notable aspects warrant discussion. Primarily, it was noted that CS primarily hinges on one core technology - the deployment of unsupervised learning algorithms. These algorithms identify commonalities among entities, enabling the clustering of borrowers exhibiting similar characteristics from varying perspectives (Jadwal et al., 2022, Machado & Karray, 2022a, Yuan et al., 2022, Moradi & Mokhatab Rafiei, 2019, Philip et al., 2018).

Among the array of clustering techniques highlighted in the literature, two approaches, namely K-means and Fuzzy C-Means (FCM), were recurrently deployed across multiple studies (Machado & Karray, 2022a, Tasgetiren et al., 2022, Yuan et al., 2022, Pandey & Shukla, 2021, Moradi & Mokhatab Rafiei, 2019, Nazari et al., 2019, Firouzabadi et al., 2018, Parmar et al., 2023). The K-means algorithm, in particular, was predominantly featured in clustering-related projects due to its versatility, accessibility, and efficiency when compared to other clustering algorithms. However, given that the efficacy of K-means and other clustering methodologies is contingent on the selection of the number of clusters, some researchers utilized established methods to streamline the selection process. For instance, the Elbow Method was adopted in a couple of studies to determine the optimal number of clusters (Jadwal et al., 2022, Machado & Karray, 2022a). Other methodologies such as the Ward Method or the Silhouette Score, which assess cluster compactness and degree of separation, were also referenced (Firouzabadi et al., 2018).

Interestingly, several studies were not strictly focused on clustering existing clients but rather introduced Hybrid Machine Learning (HML) models aiming to accomplish varying tasks (Jadwal et al., 2022, Machado & Karray, 2022a, Yuan et al., 2022, Singh et al., 2021, Lazo et al., 2020, Moradi & Mokhatab Rafiei, 2019). HML models consolidate diverse data-driven techniques and algorithms to resolve complex problems that they couldn't address independently (DOMO, 2023). For example, a few clustering-related studies initially aimed to design models predicting clients' default risk. To enhance the precision of their model, initial clusters of similar borrowers were generated to serve as a foundation upon which subsequent predictions would be executed.

Turning to the EWS-related literature, a variety of developed applications and employed techniques are evident. As Table A1 illustrates, the concept of "Early Warning Systems" is not tied to a single type of technology, but instead encompasses a multitude of practices. A prevalent implementation, discussed and utilized in several studies, revolved around creating a single risk indicator for predicting customer default risk. Within this context, two methodologies were primarily employed. The first, adopted in most of the articles, involved benchmarking numerous supervised Machine Learning algorithms to identify the most accurate and high-performing model (Wang & Zhang, 2023, Guerra et al., 2022, Petropoulos et al., 2022, Huang et al., 2022, Yang, 2022, Kita et al., 2023, Lenka et al., 2023). The second approach used statistical measurements and formulas (Zhu et al., 2021, Aytaç Emin et al., 2021, Berlinger, 2017).

Concerning the first mentioned methodology, numerous algorithms were utilized in the literature, including Logistic Regression, Support Vector Machines, Random Forest, and Decision Trees. However, XGBoost, an implementation of Gradient Boosting Decision Trees, stood out for its execution speed and accuracy, thus proving to be the most efficient algorithm (Guerra et al., 2022, Petropoulos et al., 2022, Huang et al., 2022, Yang, 2022, Solimun et al., 2023).

In addition, various EWS-related studies ventured into developing more comprehensive EWS that extended beyond merely predicting credit risk. Several articles aimed at introducing comprehensive credit risk management architectures comprising multiple modules and processes, each executing a specific function (Wangsong, 2022, Xie & Shi, 2022, Xie, 2022). Other studies sought to introduce new extensions to existing EWS, such as additional risk indicators and indexes, to augment their risk management capability (Aytaç Emin et al., 2021, Berlinger, 2017). Given this diversity of EWS applications, the field of EWS emerges as a rapidly evolving domain that lacks a unified and common framework.

In spite of the fact that modern systematic search methods have been applied very frequently over the course of the past few years, it is essential to take note of the fact that some benchmark approaches for CS and EWS have been adopted and continue to be a part of the practices that are currently in place when new methodologies are introduced. In the research that Chen et al. (2023), it was discussed that data mining is utilized to uncover comparable traits among consumers and to segment them. This and other benchmark methods (regardless of application area) have been conducted and have been cited in a number of respected journals (Andersen et al., 2009, La et al., 2009, Lapper, 2012, Prosowicz, 2024, Pagano & Jappelli, 1993, Skrastins, 2023, He & Xiong, 2012, Jorion & Zhang, 2009, Becker, 2007, Animesh et al., 2011, Albert et al., 2004, Gamba, 2022, Day, 1981, Tidhar & Eisenhardt, 2020). Similarly, EWS has been proposed and disputed in credible publications such as those indicated by citations (Berg et al., 2005, Gorr & Lee, 2015, Pekar & Burack, 1976, Li et al., 2014, Showstack, 2013, Day, 2003, Berglöf & Perotti, 1994, Altman & Loris, 1976, Pettway & Sinkey, 1980, Herrmann & Masawi, 2022, Thantharate, 2023). Although the amount of research conducted on these approaches is not as comprehensive as that conducted on CS they also need to be emphasized here.

5.3. Evaluation methods analysis

In order to accurately gauge the effectiveness of models developed in CS and EWS systems, numerous researchers have adopted various evaluation methodologies. These methodologies often span a variety of metrics and validation techniques that have been used in the relevant literature. This final subsection aims to provide a comprehensive analysis of these methodologies, shedding light on the most dominant and trending evaluation techniques.

Starting with CS models, it is noteworthy to mention that the Silhouette Score was the most common validation metric featured in the literature. Appearing in multiple studies, it was utilized to assess the quality of the created clusters (Pandey & Shukla, 2021, Nazari et al., 2019). Though the usage of multiple metrics is a norm when evaluating the performance of clustering algorithms, the Silhouette Score emerged as a common index, consistently used across a few researches.

Hybrid models, on the other hand, adopt a different approach. In these cases, the clustering of customers does not serve as the primary objective. As a result, the evaluation process is generally focused on

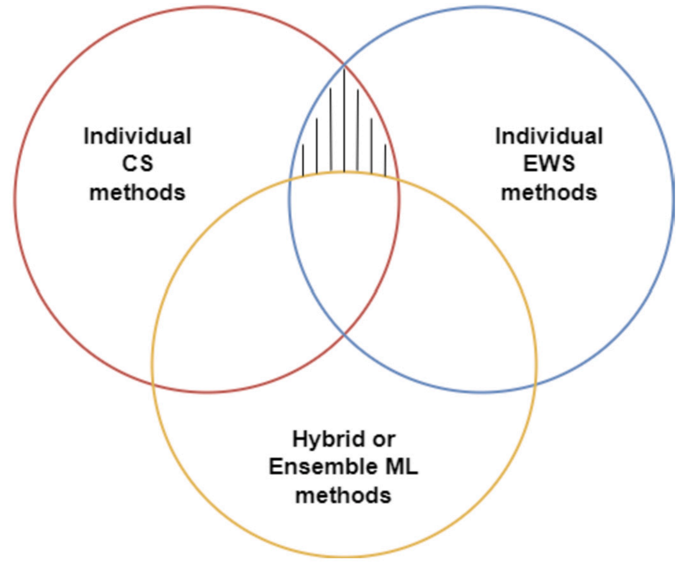

Fig. 5. The intersection between studies in CS and EWS literature.

the performance of the artefact with respect to the final outcome of the experiment. This implies that the quality validation of clusters, which serves as an intermediate step in the overall procedure, is frequently omitted or not explicitly discussed (Jadwal et al., 2022, Machado & Karray, 2022a, Yuan et al., 2022, Moradi & Mokhatab Rafiei, 2019).

When it comes to EWS-related literature, the employed evaluation methods largely depend on the nature of the application being developed. As discussed in the previous subsection, a considerable portion of the literature is dedicated to models predicting the default risk of entities based on labeled data. The evaluation of these systems typically involves computation of various metrics and application of common validation techniques used in classification-based projects. Predominant metrics include the AUC Score, Accuracy Score, and F1 Score, although other metrics like the Kolmogorov Smirnov Score and the G-Mean Score have also made appearances in several studies (Wang & Zhang, 2023, Petropoulos et al., 2022, Huang et al., 2022). Additionally, researchers have routinely relied on two primary validation techniques: train/test split and k-fold cross-validation. The latter technique usually involves the selection of the variable k based on dataset characteristics and computational resources at the researchers' disposal.

In relation to the studies focusing on more complex architectures serving as full-fledged EWS, the literature does not indicate any specific or common indicators. However, it has been observed that the effectiveness of these structures has been occasionally assessed based on qualitative analyses of the system's architecture and potential impacts that these new systems could have on existing processes (Wangsong, 2022, Zhang et al., 2019).

6. The relationship between customer segmentation methods and EWS

This section addresses the research question focused on identifying potential interconnections between Early Warning Systems (EWS) and Customer Segmentation (CS) within the existing literature. Despite the indirect nature of these connections and the lack of studies explicitly exploring the intersection of these two systems, the analysis herein unravels some intriguing and noteworthy patterns.

There was no research that was found to have a direct connection between EWS and CS, according to the systematic review. Following an exhaustive review of the Scopus database, it was discovered that there are no articles that expressly explore the interplay between these two systems. Examples of studies that focus on individual examinations of CS and EWS methodologies are depicted in Fig. 5, which is a representation of a number of studies that have been published in the academic literature (Clintworth et al., 2023, Gorle & Panigrahi, 2023, Huang et al., 2023, Lenka et al., 2023). In addition, there are studies for Customer Segmentation or Early Warning Systems frameworks that investigate hybrid or ensemble approaches. These studies include Tasgetiren et al. (2022), Madhaveelatha et al. (2023). Because of this, the next step that would make sense would be to look into the relationship that exists between these two domains.

To strengthen this claim, the earlier discussion in subsection 5 also provides some indirect indications of their potential interconnectedness. It was observed that a significant number of studies related to CS explored the development of hybrid Machine Learning models, which encompassed clustering and supervised learning algorithms for predicting customer default risk. Simultaneously, it was noted that the term "EWS" was often associated with models striving to predict similar default risks.

This correlation suggests an implicit relationship between EWS and CS technologies. Specifically, clustering algorithms, fundamental to CS systems, could play a crucial role in constructing EWS. These algorithms facilitate the detection of data patterns and risk profiles, enriching the default prediction process within EWS by integrating these insights into the financial risk estimation procedure.

Contrastingly, the literature review, as reflected in Table A1, did not reveal any studies incorporating early warning indicators within CS models. This observation suggests that this particular application remains largely untapped, warranting further exploration and experimentation. However, it is relevant to highlight how these studies are (methodologically) similar and different:

- Major Similarities:
  1. Data-Driven: Both methods heavily depend on data to operate, using historical and present data to extract insights and create forecasts or segmentations.
  2. Methods: Machine Learning and Statistical Techniques can utilize a range of algorithms to identify patterns, trends, and relationships in data.
  3. Preprocessing Requirement: It is essential for cleaning, normalizing, and transforming data to make it appropriate for analysis and modeling.
  4. Predictive Analytics: Both methods can aim to use predictive analytics for consumer segmentation to forecast future buying behaviors or segment stability, and for early warning systems to anticipate future hazards or events.
- Major Differences:
  1. Objective: Customer segmentation categorizes customers into categories with common traits for focused marketing and service delivery, while early warning systems forecast and notify about possible dangers or negative occurrences.
  2. Customer segmentation relies on customer-related data such as demographics, purchasing behavior, and engagement metrics. In contrast, early warning systems may utilize a wider variety of data sources, including economic indicators, social media signals, and operational metrics.
  3. Real-Time Processing: Early warning systems necessitate real-time or near-real-time data processing for timely risk alerts, but customer segmentation can function with batch-processed data and does not always need rapid updates.

In conclusion, the analysis herein underscores the existence of an indirect but significant link between EWS and CS systems, derived through a nuanced interpretation of the examined studies. Yet, it also highlights a noticeable gap in the literature - the integration of early warning indicators into CS systems - offering a promising direction for future research in this domain.

7. Conclusion

This research aimed to conduct a comprehensive systematic literature review on the applications of Customer Segmentation (CS) and Early Warning Systems (EWS) within the financial industry. The exploration and analysis of a diverse array of scholarly articles have culminated in invaluable insights about the prevalent trends and overarching themes that define the literature on these two sophisticated technologies. Additionally, this study endeavored to illuminate the intersectional factors that underpin the relationship between these systems and evaluate the significance of integrating EWS within CS projects.

This comprehensive and systematic approach has facilitated the addressing of the primary research questions posited at the outset:

- **Main CS and EWS methods used for credit portfolio monitoring and how they are validated:** The trends that characterize CS and EWS literature underscore a surge in published articles over recent years. The literature landscape is characterized by a broad distribution of publications across a multitude of journals, both renowned and lesser-known. A keyword analysis further reveals 'credit scoring' as a popular keyword for CS-related articles, while 'early warning systems' and 'machine learning' dominate in the context of EWS. The exploration of dominant themes centered on three primary aspects—recurring settings, data-driven techniques, and evaluation methods. In CS-related articles, a substantial focus was on retail banking clients for credit risk monitoring and online lending borrowers. In contrast, EWS-related studies spanned various industries and institutions. Concerning data-driven techniques, CS systems frequently deployed K-Means and Fuzzy C-Means, while EWS largely employed supervised ML algorithms for default risk prediction. In terms of evaluation methods, the Silhouette Score was prevalent in CS projects, whereas AUC, F1, and Accuracy scores were favored in EWS articles.
- **Relationship between CS and EWS:** The nexus between CS and EWS emerged as a crucial revelation from this research. The utility of clustering algorithms, integral to CS, was discerned within the context of predicting borrowers' default risk—an objective often associated with EWS. However, evidence suggesting the application of EWS within CS remains sparse, indicating that this particular implementation is relatively uncharted in the financial sector.

In conclusion, the analysis of customer segmentation (CS) and early warning systems (EWS) literature reveals distinct patterns and potential areas of expansion. CS studies are often set in the Peer-to-Peer (P2P) lending services context, extensively using the Lending Club dataset, with a prime focus on retail banking customers. The predominant techniques utilized are unsupervised learning algorithms such as K-means and Fuzzy C-Means (FCM).

In contrast, EWS studies are broad in their industry reach, with a notable emphasis on the Chinese market. The methodology varies greatly, encompassing single risk indicators to comprehensive EWS frameworks, and often leveraging the speedy and accurate XGBoost algorithm.

This diverse landscape points to untapped opportunities, particularly the underrepresentation of EWS applications within CS studies. This calls for potential future research that could innovate in the financial sector. By tailoring systems to specific contexts and refining methodologies, there's a significant scope to enhance the effectiveness of both CS and EWS applications.

7.1. Implications of the study

This systematic literature review provides several key implications both from a research and a practical perspective.

Research Implications: By conducting a detailed review of the existing literature, this study has identified key trends, popular techniques, and evaluation methods related to Customer Segmentation and

Early Warning Systems in the finance industry. The findings help to consolidate the existing knowledge in these areas, and more importantly, identify gaps where future research efforts can be directed. It highlights the potential for additional exploration, particularly in areas such as the integration of EWS within CS, and the expansion of the utility of EWS beyond risk mitigation.

Practical Implications: For practitioners in the finance industry, understanding the current state of research and potential applications of CS and EWS is of paramount importance. This study, therefore, provides an overview of the dominant themes in this field and may inform the design and implementation of these technologies in real-world contexts. Specifically, it points out the importance of adopting a diverse set of evaluation methods to ensure a comprehensive performance assessment, and hints towards potential innovations that could augment the effectiveness of financial decision-making systems. While it is difficult to precisely quantify the financial advantages of implementing such systems in the finance sector, it is expected that the industry can reduce losses (Machado & Karray, 2022a, 2022c, 2022b). Any improvement in minimizing customer defaults could potentially lead to significant profit maximization.

7.2. Limitations and future recommendations

While this systematic literature review has successfully offered valuable insights into the domains of Early Warning Systems (EWS) and Customer Segmentation (CS) in the financial industry, it's critical to recognize the inherent limitations that have influenced the findings. These limitations must be taken into account when interpreting the results of the study:

1. **Reliance on a single citation database and specific search terms:** The review's scope was narrowed by exclusively using the Scopus database and specific search queries to gather the literature. While this approach was adopted to streamline the literature collection process, it inevitably resulted in the omission of potentially relevant studies that were either not available on Scopus or categorized under different keywords.
2. **Application of distinct search filters:** The use of Scopus' search engine filters pertaining to article type, interest area, language, and publication year may have inadvertently prevented the inclusion of additional pertinent studies.
3. **Subjectivity in further selection:** Despite the use of specific keywords and queries to retrieve a comprehensive literature list, Scopus' recommendations included numerous articles that, based on the author's subjective evaluation, were not wholly aligned with the research objective. Consequently, these deemed inappropriate papers were excluded from the study.
4. **Limited available literature:** This systematic review also encountered limitations imposed by the literature itself. It was observed that these technologies, particularly EWS, have only recently begun to garner widespread attention, thus limiting the breadth of available research.

Taking into account the limitations encountered during the research, and building on the results yielded, several recommendations can be made to guide future research and potential implementations.

1. **Exploring the integration of EWS within CS:** As the "Evaluation Methods Analysis" subsection has indicated, none of the articles reviewed explored the incorporation of early warning triggers within customer clustering algorithms. This suggests that such an approach remains an underexplored area in the CS field. Future studies could potentially bridge this literature gap by developing systems that harmoniously blend these two technologies.
2. **Expanding the functionality of EWS:** Traditionally, EWS are designed to detect and track threats early, allowing for the implementation of timely preventive measures. However, an alternative approach could also be considered. By monitoring specific indicators of borrowers' financial health and growth, EWS could be used to identify positive changes in financial behavior. Such insights could signal promising lending opportunities, expanding the utility of EWS beyond risk mitigation.

In conclusion, the scope for innovation and expansion within the fields of Customer Segmentation and Early Warning Systems is significant. By exploring underrepresented applications and adopting novel approaches to existing technologies, future research can continue to enhance our understanding of these advanced systems, ultimately contributing to more effective and comprehensive financial decision-making processes.

CRediT authorship contribution statement

Alessandra Amato: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Joerg R. Osterrieder:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Marcos R. Machado:** Writing – review & editing, Supervision, Project administration, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work has been supported by several institutions, each of which has provided vital resources and expertise to the research project. Firstly, we acknowledge the COST Action CA19130 and COST Action CA21163, under the auspices of the European Cooperation in Science and Technology (COST). COST Actions provide networking opportunities for researchers across Europe, fostering scientific exchange and innovation.

We would like to express our gratitude to the Swiss National Science Foundation for its financial support across multiple projects. This includes the project on Mathematics and Fintech (IZCNZ0-174853), which focuses on the digital transformation of the Finance industry. We also appreciate the funding for the project on Anomaly and Fraud Detection in Blockchain Networks (IZSEZ0-211195), and for the project on Narrative Digital Finance: a tale of structural breaks, bubbles & market narratives (IZCOZ0-213370). In addition, our research has benefited from funding from the European Union's Horizon 2020 research and innovation program under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA). This grant was provided for the FIN-TECH project, a training programme aimed at promoting compliance with financial supervision and technology.

We gratefully acknowledge the support of the Marie Skłodowska-Curie Actions under the European Union's Horizon Europe research and innovation program for the Industrial Doctoral Network on Digital Finance, acronym: DIGITAL, Project No. 101119635. Their significant contribution has been instrumental in advancing our research and fostering collaboration within the digital finance field across Europe. Next to that, we also would like to acknowledge that this research has received funding from the European Union's HORIZON Research and Innovation Actions programme under grant agreement number 101138473, related to the project "IS2H4C - From Industrial Symbiosis to Hub for Circularity".

Lastly, we acknowledge the cooperative relationship between the ING Group and the University of Twente. This partnership, centered on advancing Artificial Intelligence in Finance in the Netherlands and beyond, has been of great value to our research.

These partnerships and funding sources have greatly contributed to our ability to conduct rigorous and impactful research. Our findings are our own and do not necessarily represent the views of the supporting institutions.

Appendix A

Table A1
Table reporting all the articles that have been examined to conduct the research and highlighting their main features.

Topic	Author	Settings	Main Purpose	Data-Driven Technique[a]	Metrics for evaluation[b]
CS	Kaminskyi et al. (2022)	Online Lending	Customer Segmentation using a scoring approach and creation of an indicator of expected return of a borrower to determine a marketing solution	Whale curve for segmentation based on cumulative income, LR for creating scoring construction and SoftMax for generating vectors of probabilities of belonging to segments	–
CS	Jadwal et al. (2022)	Peer-to-Peer Lending	Customer Segmentation combined with SMOTE oversampling to reduce class imbalance and improve default prediction	Spectral clustering using the Elbow Method, benchmarking various SMOTE oversampling techniques on the clusters presenting a major of minor class instances, implementation of LR, SVM and KNN to test the effectiveness of the oversampling algorithms	G-Mean and F1 Score
CS	Machado and Karray (2022a)	Peer-to-Peer Lending	Prediction of Customer's Risk-Adjusted Revenue (RAR) integrating Customer Segmentation	K-means using Elbow Method and DBSCAN for customer clustering, six regressors (AB, GB, DT, RF, SVM, and ANN) to predict RAR	EV, R^2, MAE, MSE, MedAE
CS	Tasgetiren et al. (2022)	Banks	Hybrid distributed software architecture to segment customers and predict loan usage tendency	XGBoost and LightGBM for supervised learning, K-means, Bisecting K-means and Gaussian Mixture Model for unsupervised	Latency Test, Accuracy Test and Scalability Test for supervised Machine Learning workflows, Purity and Entropy metrics for unsupervised workflows
CS	Yuan et al. (2022)	Banks and Lending Institutions	Default Prediction integrating customer clustering	K-Means Clustering for segmenting customers, SVDD for one-classification	AUC, G-Mean and Type-II Error
CS	Pandey and Shukla (2021)	Financial Institutions	Stratified Remainder Linear Systematic Sampling Extension (SRSE) to improve computational efficiency of risk clustering algorithms	Benchmarking the SSE-based clustering approach to the classical partitional K-means and K-means ++	Davies Bouldin score, Silhouette coefficient, Scattering Density between clusters Validity, Scattering Distance Validity and CPU Time
CS	Singh et al. (2021)	Financial Instances	Classification of good and bad credit	Multi-Level Classification using 4 base classifiers (NN, KNN, SVM and RF), variation of Particle Swam Optimization algorithm for clustering the training dataset after the first classification to assign weights to the classifiers in different spacial regions	H measure, Precision, AUC, Recall, F1 Score and Accuracy
CS	Lazo et al. (2020)	Financing companies in the agrobusiness sector	Prediction of the probability of default of Chilean farmers	ClustOfVar algorithm for clustering features, LR, NN and for predicting the PD	AUC for the prediction, Mean Decrease Gini (MDG) and Mean Decrease Accuracy (MDA) to measure the importance of each variable
CS	Moradi and Mokhatab Rafiei (2019)	Banks	Predicting banks credit risk level	Fuzzy C-Means (FCM) for clustering the clients, ANFIS to predict customers risk level, application of Fuzzy Interference System (FIS) on medium risk customers to identify "too risky" borrowers	Degree of sensitivity and degree of diagnosis
CS	Nazari et al. (2019)	Banks and Credit Institutions	Customer Segmentation and credit scoring	Benchmarking K-Means, FCM and Sub-clustering techniques	LIFT and Silhouette Scores
CS	Philip et al. (2018)	Financial Institutions	Clustering clients over time based on the repayment behavior	Benchmarking normal and Partially Constrained Hidden Markov Models (PC-HMM) to cluster time series data	Measuring how the index of bank branch matches the clusters generated
CS	Firouzabadi et al. (2018)	Banks	Customer Segmentation for the optimal allocation of bank services	K-Means for clustering, suing Ward Method and Silhouette Score for optimal number of clusters	Distribution analysis of the variables within clusters
CS	Luthfi and Wibowo (2017)	Financial Institutions	Predicting loan payments using ANFIS	ANFIS	RMSE
CS	Kita et al. (2023)	Customer Behavior in Banking	Predicting consumer behavior based on a two-step clustering method	Decision Tree models	Statistical analysis, accuracy, AUC, and F-1 score

Table A1 (*continued*)

Topic	Author	Settings	Main Purpose	Data-Driven Technique[a]	Metrics for evaluation[b]
CS	Parmar et al. (2023)	Enterprises	rebalancing data and cluster examples using Fuzzy clustering algorithms	Fuzzy clustering methods	Accuracy, precision, recall, and f1-score metrics
CS	Lenka et al. (2023)	Banks and Financial Institutions	Analyzing a novel adaptive clustering algorithm to handle imbalanced credit-scoring datasets	Semi-supervised clustering algorithms	F1-score, AUC, and G-mean
CS	Solimun et al. (2023)	Loan/Mortgage settings	Integrating cluster analysis with path analysis to model compliance behavior for paying the House Ownership Loan	Clustering and path analysis models	Statistical analysis and tests
CS	Liu et al. (2023)	Banks and financial institutions	Test a Multi-Level Default Risk Rating (MLDRR) strategy based on a heterogeneous ensemble learning method	Clustering, Supervised ML classifiers, and Statistical analysis	Statistical tests and inference, F1, accuracy
CS	Huang et al. (2023)	Enterprises	Forecast financial distress based on sentiment analysis	Annual reports sentiment analysis models and clustering algorithms	Sentiment scores and accuracy
CS	Boyapati and Aygun (2023)	Banks and financial institutions	Analyze the use of a graph-based variable clustering (GVC) method as a filter based approach to select prominent features while retaining as much variance as possible	(Un)Supervised ML models	Processing time and accuracy
CS	Gorle and Panigrahi (2023)	Loan settings - Fraud detection	Explore the use of a semi-supervised outlier score-based Anti-Fraud model to identify the loan applicant as a genuine or fraudulent debtor	(Un)Supervised ML Models	Accuracy, precision and recall
CS	Vlahavas et al. (2024)	Cryptocurrecies - Bitcoin	Testing a set of features that can be extracted from transaction data to group transactions to support decision-making process	k-Means and multiple feature engineering methods	Characterize clusters
CS	Chen et al. (2023)	Online Loan Platforms	Proposing a credit rating system for Online Loans in China and testing clustering algorithms to cluster company's credit score rating	Statistics and k-Means	Accuracy
CS	Han et al. (2023)	Loan Settings	To analyze the viability on using quantitative methods to process and examine short loan audit texts	Fuzzy clustering analysis (FCA) and Clustering algorithms	Accuracy and Statistical tests
CS	Ahelegbey and Giudici (2023)	Loan Settings (P2P)	Improve measurement of credit scoring by applying factor clustering	k-Means	Characterization of clustering and similarity scores
CS	Yin et al. (2023)	Loan Settings (P2P)	Predicting credit default probabilities for P2P lending by implementing ML techniques	(Un)Supervised ML models	Accuracy, precision, and recall
CS	Mayorova et al. (2023)	Banking and Loan industry	Assessing the level of financial inclusion of households in the regions of Ukraine in the market of bank loans in the pre-war period	Clustering methods	Clusters characterization and statistical tests
EWS	Wang and Zhang (2023)	Chinese manufacturing companies	Predicting high credit risk companies	Two-stage ensemble model: the first stage used Grey Relational Analysis to select relevant indicators, the second stage implemented the Bagging Method to integrate 5 CNN models to make the prediction based on 5-fold cross-validation	Accuracy, Recall, Precision, F1 Score, G-Mean
EWS	Guerra et al. (2022)	Portuguese bank	Classification of banks' risk level	RF Classifier for feature selection, benchmarking of LR, SVM Classifier, NB Classifier, RF Classifier and XGBoost Classifier using train-test split, 10-fold cross validation and TPOT	F1 Score and Confusion Matrix
EWS	Petropoulos et al. (2022)	Sovereigns and governments	Creation of a sovereign rating system based on their risk of default	Benchmarking LR, SVM, NN, RF, CIF and XGBoost using train-test split, calibration to implement the credit rating system	AUC, Kolmogorov-Smirnov test, Youden Score, Negative Likelihood Ratio, Geometric Mean Balanced Accuracy for validating the prediction performance, SSE and Brier Score to validate the calibration
EWS	Wangsong (2022)	Internet Finance of Chinese banks	Use multimedia technology to design a bank customer default risk management system that measures and controls credit risk	–	Test performance of the system and users survey feedback

(continued on next page)

Table A1 (*continued*)

Topic	Author	Settings	Main Purpose	Data-Driven Technique[a]	Metrics for evaluation[b]
EWS	Xie and Shi (2022)	Chinese companies with regards to Internet Finance	Creation of a Internet Financial Risk Control model using Big Data	Criteria Importance Through Intercriteria Correlation (CRITIC) method to determine the weight of the risks indexes and mathematical expression to calculate the probability of the financial risk	Error Rate, Accuracy, Control time
EWS	Han (2022)	Governments	Creation of Economic Data Management System	2 Back Propagation (2BP) Neural Network	–
EWS	Yin et al. (2022)	Supply Chain Finance	Creation of a Risk Early Warning Index System to predict the default risk	PCA for selection of Risk Early Warning Indicators, CNN for predicting the risk index value	Accuracy and Loss Function
EWS	Xie (2022)	Supply-chain Financing	Creation of a blockchain-based financial risk prevention system	Evolutionary game model to measure the endorsement fiduciary relationship between SMEs and major enterprises	–
EWS	Huang et al. (2022)	Chinese firms	Predicting financial distress of companies using textual sentiment of annual reports	word2vec to create word vectors, NB, SVM, DT, KNN, CNN and LSTM for generating textual sentiment using train-test split, benchmarking five different classifiers (LR, NN, LS-SVM, RF and XGBoost) based on out-of-time and out-of-sample predictions using train/test and k-fold cross validation	AUC, Kolmogorov–Smirnov statistic, Brier score, precision, recall, F1 score, and total accuracy for out-of-time prediction
EWS	Xu et al. (2022)	Countries	Creation of a Machine Learning-based risk awareness model for financial crisis prediction based on a number of risk factors	Design, Solo and Easy Classification techniques	Accuracy Ratio
EWS	Tong and Tong (2022)	Enterprises	Predicting the Cash Flow risk	DT	Entropy, Split information and Gain ratio
EWS	Yang (2022)	Online Lending	Creation of Credit Risk Assessment model that predicts default risk	Benchmarking Random Forest, XGBoost and XGBoost Deep Forest	AUC
EWS	Jacobs (2021)	Publicly rated US companies	Predicting the Point-In-Time (PIT) and Through-The-Cycle (TTC) Probability of Default (PD) in one and three years horizons	LR	AUC, Hosmer-Lemeshow test, Akaike Information Criterion (AIC), Singular Value Decomposition and Factor Contribution (for individual variables)
EWS	Zhu et al. (2021)	IoT companies	Creation of a Risk Early Warning System composed of a Z-Score Model analysis to predict corporate finance and bankruptcy risk	Z-Score Model	–
EWS	Aytaç Emin et al. (2021)	Islamic Banks	Creation of a new banking fragility index to improve the predictive power of an Early Warning System	Mathematical Formula for calculating the banking fragility index the	Predictive Power
EWS	Zhang et al. (2019)	Accounting departments of companies	Improving the company's financial model and evaluating the effectiveness of the internal control of the financial reporting system	–	Qualitative analysis
EWS	Pompella and Dicanio (2017)	Publicly rated banks	Creation of an EWS that uses an accounting-based approach to identify high and low risk banks in order to test the validity of external ratings	Principal Commponent-Mahalanobis (PC-M) method	ROC curve
EWS	Berlinger (2017)	Interbank lending market	Introduction of a new indicator, called Implicit Rating (IR), for EWS	Mathematical formula for calculating the IR score	–
EWS	Hacibedel and Qu (2023)	Enterprises	Understanding and predicting customers distress	ML models for EWS	Feature importance analysis and accuracy measurement
EWS	Ramesh and Jeyakarthic (2023)	Banks and financial institutions	Outlier detection (EWS) for financial credit score prediction	Fuzzy Support Vector Machines	Accuracy, F1, precision, recall
EWS	Clintworth et al. (2023)	Banks and financial institutions	Designing, developing and testing of a novel ML methodology which integrates predictor evaluation and missing data analysis into the distress prediction process (EWS)	(Un)Supervised ML models	Accuracy and statistical tests
EWS	Madhaveelatha et al. (2023)	Enterprises	To test the efficiency of a hybrid ML models for filtering out superfluous data maintaining important information in datasets	(Un)Supervised ML models	Accuracy, AUC, precision, recall

[a] Note: KNN refers to K-Nearest Neighbor, SVM is Support Vector Machine, CNN is Convolution Neural Networks, ANN is Artificial Neural Networks, AB is Adaboost, RF is Random Forest, GB is Gradient Boosting, DT is Decision Tree, NB is Naive Bayes, LR is Logistic Regression, SVDD is Support Vector Domain Description, ANFIS is Adaptive Neuro-Fizzy Inference Systems, CIF is Conditional Inference Trees, LSTM is Long Short-Term Memory.

[b] Note: MSE is Mean Squared Error, RMSE is the Root Mean Squared Error, MAE is the Mean Absolute Error, ROC/AUC refers to the Area Under the ROC Curve, MedAE is Median Absolute Error.

References

Ahelegbey, D. F., & Giudici, P. (2023). Credit scoring for peer-to-peer lending. *Risks*, *11*. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85166413114&doi=10.3390%2frisks11070123&partnerID=40&md5=fcc8a1fcf1db8d6b0d4bc48ab6748ddc. Cited by: 0; All Open Access, Gold Open Access.

Albert, T. C., Goes, P. B., & Gupta, A. (2004). Gist: A model for design and management of content and interactivity of customer-centric web sites. *MIS Quarterly*, 161–182.

Altman, E. I., & Loris, B. (1976). A financial early warning system for over-the-counter broker-dealers. *The Journal of Finance*, *31*, 1201–1217.

Andersen, P. H., Christensen, P. R., & Damgaard, T. (2009). Diverging expectations in buyer–seller relationships: Institutional contexts and relationship norms. *Industrial Marketing Management*, *38*, 814–824.

Animesh, A., Viswanathan, S., & Agarwal, R. (2011). Competing "creatively" in sponsored search markets: The effect of rank, differentiation strategy, and competition on performance. *Information Systems Research*, *22*, 153–169.

Aytaç Emin, A., Dalgıç, B., & Azrak, T. (2021). Constructing a banking fragility index for Islamic banks: Definition impact on the predictive power of an early warning system. *Applied Economics Letters*, *28*, 1589–1593. https://doi.org/10.1080/13504851.2020.1834497.

Bach, M. P., Juković, S., Dumičić, K., & Šarlija, N. (2013). Business client segmentation in banking using self-organizing maps. *South East European Journal of Economics and Business*, *8*, 32–41.

Becker, B. (2007). Geographical segmentation of us capital markets. *Journal of Financial Economics*, *85*, 151–178.

Berg, A., Borensztein, E., & Pattillo, C. (2005). Assessing early warning systems: How have they worked in practice? *IMF Staff Papers*, *52*, 462–502.

Berglöf, E., & Perotti, E. (1994). The governance structure of the Japanese financial keiretsu. *Journal of Financial Economics*, *36*, 259–284.

Berlinger, E. (2017). Implicit rating: A potential new method to alert crisis on the interbank lending market. *Finance Research Letters*, *21*, 277–283. https://doi.org/10.1016/j.frl.2016.11.010.

Boyapati, M., & Aygun, R. (2023). Default prediction on commercial credit big data using graph-based variable clustering. (p. 139–142), https://www.scopus.com/inward/record.uri?eid=2-s2.0-85151567055&doi=10.1109%2fICSC56153.2023.00029&partnerID=40&md5=dab24b0d1fdd41a40c7db7d0eb87fd0c. cited by: 0.

Chen, R., Wang, S., Zhu, Z., Yu, J., & Dang, C. (2023). Credit ratings of Chinese online loan platforms based on factor scores and k-means clustering algorithm. *Journal of Management Science and Engineering*, *8*, 287–304. https://doi.org/10.1016/j.jmse.2022.12.003. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85153797809&doi=10.1016%2fj.jmse.2022.12.003&partnerID=40&md5=2c0104a52e0588fd4941ae8a84a21631. Cited by: 0; All Open Access, Gold Open Access.

Clintworth, M., Lyridis, D., & Boulougouris, E. (2023). Financial risk assessment in shipping: A holistic machine learning based methodology. *Maritime Economics and Logistics*, *25*, 90–121. https://doi.org/10.1057/s41278-020-00183-2. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85098792497&doi=10.1057%2fs41278-020-00183-2&partnerID=40&md5=5bcec820bb4f99bb3c5ab420236183a7. Cited by: 4; All Open Access, Green Open Access.

Day, B. A. (2003). Early warning system scores and response times: An audit. *Nursing in Critical Care*, *8*, 156–164.

Day, G. S. (1981). Strategic market analysis and definition: An integrated approach. *Strategic Management Journal*, *2*, 281–299.

DOMO (2023). Data science and ml platform: What is hybrid machine learning? https://www.domo.com/glossary/what-is-hybrid-machine-learning.

ECB (2017). European Central Bank (ECB) - guidance to banks on non-performing loans. https://www.bankingsupervision.europa.eu/ecb/pub/pdf/guidance_on_npl.en.pdf.

Firouzabadi, S. M. A. K., Taghavifard, M. T., Sajjadi, S. K., & Soufi, J. B. (2018). A multi-objective optimisation model for assignment of service to bank customers by using data mining and simulation. *International Journal of Electronic Customer Relationship Management*, *11*, 237–255. https://doi.org/10.1504/IJECRM.2018.093766.

Fujii, M., Sakaji, H., Masuyama, S., & Sasaki, H. (2022). Extraction and classification of risk-related sentences from securities reports. *International Journal of Information Management Data Insights*, *2*, Article 100096. https://doi.org/10.1016/j.jjimei.2022.100096. https://www.sciencedirect.com/science/article/pii/S2667096822000398.

Gamba, D. (2022). Servitized customer segmentation model for smes. In *Academy of management proceedings: Vol. 10510 (p. 18212)*. NY: Academy of Management Briarcliff Manor. 2022.

Gorle, V. L. N., & Panigrahi, S. (2023). A semi-supervised anti-fraud model based on integrated xgboost and bigru with self-attention network: An application to Internet loan fraud detection. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-17681-z. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85179697758&doi=10.1007%2fs11042-023-17681-z&partnerID=40&md5=048ae0cd3fe0de14c4722e0937dfe1b8. Cited by: 0.

Gorr, W. L., & Lee, Y. (2015). Early warning system for temporary crime hot spots. *Journal of Quantitative Criminology*, *31*, 25–47.

Guerra, P., Castelli, M., & Côrte-Real, N. (2022). Machine learning for liquidity risk modelling: A supervisory perspective. *Economic Analysis and Policy*, *74*, 175–187. https://doi.org/10.1016/j.eap.2022.02.001.

Hacibedel, B., & Qu, R. (2023). Understanding and predicting systemic corporate distress: A machine-learning approach. *Journal of Credit Risk*, *19*, 79–116. https://doi.org/10.21314/JCR.2023.006. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85172916190&doi=10.21314%2fJCR.2023.006&partnerID=40&md5=c91e6dd6d813f86c3364532ea132a7ca. Cited by: 0.

Han, L., Liu, Z., Qiang, J., & Zhang, Z. (2023). Fuzzy clustering analysis for the loan audit short texts. *Knowledge and Information Systems*, *65*, 5331–5351. https://doi.org/10.1007/s10115-023-01943-1. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85165553165&doi=10.1007%2fs10115-023-01943-1&partnerID=40&md5=314d3e01493a8dd53d8e0847f06cbefc. Cited by: 0.

Han, X. (2022). Construction of economic data management system based on bp neural network. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/9036917.

He, Z., & Xiong, W. (2012). Rollover risk and credit risk. *The Journal of Finance*, *67*, 391–430.

Herrmann, H., & Masawi, B. (2022). Three and a half decades of artificial intelligence in banking, financial services, and insurance: A systematic evolutionary review. *Strategic Change*, *31*, 549–569.

Hu, X., Liu, A., Li, X., Dai, Y., & Nakao, M. (2023). Explainable ai for customer segmentation in product development. *CIRP Annals*, *72*, 89–92.

Huang, B., Yao, X., Luo, Y., & Li, J. (2022). Improving financial distress prediction using textual sentiment of annual reports. *Annals of Operations Research*, 1–28. https://doi.org/10.1007/s10479-022-04633-3.

Huang, B., Yao, X., Luo, Y., & Li, J. (2023). Improving financial distress prediction using textual sentiment of annual reports. *Annals of Operations Research*, *330*, 457–484. https://doi.org/10.1007/s10479-022-04633-3. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85126326187&doi=10.1007%2fs10479-022-04633-3&partnerID=40&md5=d372b32197e09cb53d61107419575049. Cited by: 5.

Jacobs, M. (2021). Validation of corporate probability of default models considering alternative use cases. *International Journal of Financial Studies*, *9*, 63. https://doi.org/10.3390/ijfs9040063.

Jadwal, P. K., Jain, S., Pathak, S., & Agarwal, B. (2022). Improved resampling algorithm through a modified oversampling approach based on spectral clustering and smote. *Microsystem Technologies*, 1–9. https://doi.org/10.1007/s00542-022-05287-8.

Jorion, P., & Zhang, G. (2009). Credit contagion from counterparty risk. *The Journal of Finance*, *64*, 2053–2087.

Kaminskyi, A., Nehrey, M., Babenko, V., & Zimon, G. (2022). Model of optimizing correspondence risk-return marketing for short-term lending. *Journal of Risk and Financial Management*, *15*, 583. https://doi.org/10.3390/jrfm15120583.

Kansal, T., Bahuguna, S., Singh, V., & Choudhury, T. (2018). Customer segmentation using k-means clustering. In *2018 international conference on computational techniques, electronics and mechanical systems (CTEMS)* (pp. 135–139). IEEE.

Kashwan, K. R., & Velu, C. (2013). Customer segmentation using clustering and data mining techniques. *International Journal of Computer Theory and Engineering*, *5*, 856.

Kita, P., Žambochová, M., Maciejewski, G., Čvirik, M., & Mazalánová, V. K. (2023). Changes in the culture of consumption during covid-19: A decision-tree model. *Cultural Management: Science and Education*, *7*, 85–101. https://doi.org/10.30819/cmse.7-1.06. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85172283419&doi=10.30819%2fcmse.7-1.06&partnerID=40&md5=94b1b0883c9d2947e1be184981ac8534. Cited by: 0.

Klopotan, I., Zoroja, J., & Meško, M. (2018). Early warning system in business, finance, and economics: Bibliometric and topic analysis. *International Journal of Engineering Business Management*, *10*, Article 1847979018797013. https://doi.org/10.1177/1847979018797013.

Kushwaha, A. K., Kar, A. K., & Dwivedi, Y. K. (2021). Applications of big data in emerging management disciplines: A literature review using text mining. *International Journal of Information Management Data Insights*, *1*, Article 100017.

La, V., Patterson, P., & Styles, C. (2009). Client-perceived performance and value in professional b2b services: An international perspective. *Journal of International Business Studies*, *40*, 274–300.

Lapper, R. (2012). Opinion: Why segmentation is vital to success. *Financial Times*. https://www.ft.com/content/04fd1c9a-b47d-11e1-bb2e-00144feabdc0.

Lazo, D., Calabrese, R., & Bravo, C. (2020). The effects of customer segmentation, borrower behaviors and analytical methods on the performance of credit scoring models in the agribusiness sector. *Journal of Credit Risk*, *16*. https://doi.org/10.21314/JCR.2020.272.

Lenka, S. R., Bisoy, S. K., & Priyadarshini, R. (2023). A-rdbote: An improved oversampling technique for imbalanced credit-scoring datasets. *Risk Management*, *25*. https://doi.org/10.1057/s41283-023-00128-y. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85168504280&doi=10.1057%2fs41283-023-00128-y&partnerID=40&md5=513b1e10d34d9c8c3b8c1cb9b4bbe0a6. Cited by: 0.

Li, J., Qin, Y., Yi, D., Li, Y., & Shen, Y. (2014). Feature selection for support vector machine in the study of financial early warning system. *Quality and Reliability Engineering International*, *30*, 867–877.

Liu, S., Wei, G., Wu, S., & Sun, Y. (2023). An ensemble learning based strategy for customer subdivision and credit risk characterization. *Tehnicki Vjesnik*, *30*, 426–433. https://doi.org/10.17559/TV-20221220085239. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85149295929&doi=10.17559%2fTV-20221220085239&partnerID=40&md5=6e9153e2c270ba8fbde4e33b71c9d9a3. Cited by: 0; All Open Access, Gold Open Access.

Luthfi, E. T., & Wibowo, F. W. (2017). Loan payment prediction using adaptive neuro fuzzy inference system. *International Journal of Simulation: Systems, Science & Technology*, *18*, 9.1–9.6. https://doi.org/10.5013/IJSSST.a.18.04.09.

Machado, M. R., & Karray, S. (2022a). Applying hybrid machine learning algorithms to assess customer risk-adjusted revenue in the financial industry. *Electronic Commerce Research and Applications*, *56*, Article 101202. https://doi.org/10.1016/j.elerap.2022.101202.

Machado, M. R., & Karray, S. (2022b). Assessing credit risk of commercial customers using hybrid machine learning algorithms. *Expert Systems with Applications*, *200*, Article 116889.

Machado, M. R., & Karray, S. (2022c). Integrating customer portfolio theory and the multiple sources of risk approaches to model risk-adjusted revenue. *IFAC-papersonline*, *55*, 356–363.

Madhaveelatha, A., Varaprasad, K., & Pydala, B. (2023). Classification model for identification of Internet loan frauds using pca with ensemble method. In *Lecture notes in networks and systems: Vol. 649 LNNS* (pp. 486–495). https://www.scopus.com/inward/record.uri?eid=2-s2.0-85152540308&doi=10.1007%2f978-3-031-27499-2_46&partnerID=40&md5=de1a039d3e4ec5f848a75a788ee71ee6. Cited by: 0.

Mayorova, T., Tymoshenko, I., Urvantseva, S., Chernyak, R., & Lutsiv, P. (2023). Cluster approach in assessing the pre-war level of financial inclusion of population from different regions of Ukraine in the market of bank loan. *Financial and Credit Activity: Problems of Theory and Practice*, *1*, 64–76. https://doi.org/10.55643/fcaptp.1.48.2023.3975. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85164670353&doi=10.55643%2ffcaptp.1.48.2023.3975&partnerID=40&md5=d779566f83d17454d27166bdd9699b55. Cited by: 0; All Open Access, Gold Open Access.

Mihova, V., & Pavlov, V. (2018). A customer segmentation approach in commercial banks. In *AIP conference proceedings: Vol. 2025*. AIP Publishing LLC (p. 030003), https://doi.org/10.1063/1.5064881.

Moradi, S., & Mokhatab Rafiei, F. (2019). A dynamic credit risk assessment model with data mining techniques: Evidence from Iranian banks. *Financial Innovation*, *5*, 1–27. https://doi.org/10.1186/s40854-019-0121-9.

Mousaeirad, S. (2020). Intelligent vector-based customer segmentation in the banking industry. arXiv preprint, arXiv:2012.11876.

Nazari, A., Mehregan, M., & Tehrani, R. (2019). Credit scoring of bank depositor with clustering techniques for supply chain finance. *International Journal of Supply Chain Management*, *8*, 374–383.

Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, *36*, 2592–2602.

Pagano, M., & Jappelli, T. (1993). Information sharing in credit markets. *The Journal of Finance*, *48*, 1693–1718.

Pandey, K. K., & Shukla, D. (2021). Stratified remainder linear systematic sampling based clustering model for loan risk detection in big data mining. *Reliability: Theory & Applications*, *16*, 239–257.

Pandey, M. K., Mittal, M., & Subbiah, K. (2021). Optimal balancing & efficient feature ranking approach to minimize credit risk. *International Journal of Information Management Data Insights*, *1*, Article 100037. https://doi.org/10.1016/j.jjimei.2021.100037. https://www.sciencedirect.com/science/article/pii/S2667096821000306.

Parmar, G., Gupta, R., Bhatt, T., Sahani, G., Panchal, B. Y., & Patel, H. (2023). Data re-balancing using fuzzy clustering and smot mechani (pp. 1714–1718). https://www.scopus.com/inward/record.uri?eid=2-s2.0-85168315249&doi=10.1109%2fICESC57686.2023.10192964&partnerID=40&md5=7c349afeb584f27e59eac366e8e023ed. cited by: 0.

Pekar Jr, P. P., & Burack, E. H. (1976). Management control of strategic plans through adaptive techniques. *Academy of Management Journal*, *19*, 79–97.

Petropoulos, A., Siakoulis, V., & Stavroulakis, E. (2022). Towards an early warning system for sovereign defaults leveraging on machine learning methodologies. *Intelligent Systems in Accounting, Finance & Management*, *29*, 118–129. https://doi.org/10.1002/isaf.1516.

Pettway, R. H., & Sinkey, J. F. (1980). Establishing on-site bank examination priorities: An early-warning system using accounting and market information. *The Journal of Finance*, *35*, 137–150.

Philip, D. J., Sudarsanam, N., & Ravindran, B. (2018). Improved insights on financial health through partially constrained hidden Markov model clustering on loan repayment data. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *49*, 98–113. https://doi.org/10.1145/3242734.3242741.

Pompella, M., & Dicanio, A. (2017). Ratings based inference and credit risk: Detecting likely-to-fail banks with the pc-Mahalanobis method. *Economic Modelling*, *67*, 34–44. https://doi.org/10.1016/j.econmod.2016.08.023.

Prosowicz, G. (2024). Customer-centricity in wealth management. https://www.ft.com/partnercontent/comarch/customer-centricity-in-wealth-management.html.

Raiter, O. (2021). Segmentation of bank consumers for artificial intelligence marketing. *International Journal of Contemporary Financial Issues*, *1*, 39–54. https://doi.org/10.17613/q0h8-m266.

Ram, J., Zhang, C., & Koronios, A. (2016). The implications of big data analytics on business intelligence: A qualitative study in China. *Procedia Computer Science*, *87*, 221–226. https://doi.org/10.1016/j.procs.2016.05.152.

Ramesh, R., & Jeyakarthic, M. (2023). Fuzzy support vector machine based outlier detection for financial credit score prediction system. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *14*, 60–73. https://doi.org/10.58346/JOWUA.2023.I4.005. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85181493981&doi=10.58346%2fJOWUA.2023.I4.005&partnerID=40&md5=4653fb381e87b39a0bce0ce30a06a06c. Cited by: 0; All Open Access, Bronze Open Access.

S, V. P., Chakraborty, A., & Kar, A. K. (2024). How to undertake an impactful literature review: Understanding review approaches and guidelines for high-impact systematic literature reviews. *South Asian Journal of Business and Management Cases*. https://doi.org/10.1177/22779779241227654.

Sasaki, H., Fujii, M., Sakaji, H., & Masuyama, S. (2024). Enhancing risk identification with gnn: Edge classification in risk causality from securities reports. *International Journal of Information Management Data Insights*, *4*, Article 100217. https://doi.org/10.1016/j.jjimei.2024.100217. https://www.sciencedirect.com/science/article/pii/S2667096824000065.

Showstack, R. (2013). Concern about abrupt climate changes prompts call for early warning system.

Singh, I., Kumar, N., Srinivasa, K., Maini, S., Ahuja, U., & Jain, S. (2021). A multi-level classification and modified pso clustering based ensemble approach for credit scoring. *Applied Soft Computing*, *111*, Article 107687. https://doi.org/10.1016/j.asoc.2021.107687.

Skrastins, J. (2023). Barter credit: Warehouses as a contracting technology. *The Journal of Finance*, *78*, 2009–2047.

Solimun, W. N. W. S., Fernandes, A. A. R., Kartikasari, D. P., & Hutahayan, B. (2023). Cluster analysis with various combination of distance and linkage for modeling dummy variable path analysis. *Advanced Mathematical Models and Applications*, *8*, 565–589. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85184397527&partnerID=40&md5=11a7c08047e1f993e2bcde7cb1e55d4c.

Talaat, F. M., Aljadani, A., Alharthi, B., Farsi, M. A., Badawy, M., & Elhosseini, M. (2023). A mathematical model for customer segmentation leveraging deep learning, explainable ai, and rfm analysis in targeted marketing. *Mathematics*, *11*, 3930.

Tasgetiren, N., Tigrak, U., Bozan, E., Gul, G., Demirci, E., Saribiyik, H., & Aktas, M. S. (2022). On the distributed software architecture of a data analysis workflow: A case study. *Concurrency and Computation: Practice and Experience*, *34*, Article e6522. https://doi.org/10.1002/cpe.6522.

Thantharate, P. (2023). Intelligentmonitor: Empowering devops environments with advanced monitoring and observability. In *2023 international conference on information technology (ICIT)* (pp. 800–805). IEEE.

Tidhar, R., & Eisenhardt, K. M. (2020). Get rich or die trying... finding revenue model fit using machine learning and multiple cases. *Strategic Management Journal*, *41*, 1245–1273.

Tong, L., & Tong, G. (2022). A novel financial risk early warning strategy based on decision tree algorithm. *Scientific Programming*, *2022*, 1–10. https://doi.org/10.1155/2022/4648427.

Vlahavas, G., Karasavvas, K., & Vakali, A. (2024). Unsupervised clustering of bitcoin transactions. *Financial Innovation*, *10*. https://doi.org/10.1186/s40854-023-00525-y. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85182634332&doi=10.1186%2fs40854-023-00525-y&partnerID=40&md5=8e37d97743c1732158817abee0d44eb1. Cited by: 0; All Open Access, Gold Open Access.

Wang, L., & Zhang, W. (2023). A qualitatively analyzable two-stage ensemble model based on machine learning for credit risk early warning: Evidence from Chinese manufacturing companies. *Information Processing & Management*, *60*, Article 103267. https://doi.org/10.1016/j.ipm.2023.103267.

Wangsong, X. (2022). The default risk of bank customers based on embedded microprocessor wireless communication under the internet finance background. *Mobile Information Systems*, *2022*. https://doi.org/10.1155/2022/8019033.

Xie, H., & Shi, Y. (2022). A big data technique for internet financial risk control. *Mobile Information Systems*, *2022*. https://doi.org/10.1155/2022/9549868.

Xie, W. (2022). Study on enterprise financial risk prevention and early warning system based on blockchain technology. *Mobile Information Systems*, *2022*. https://doi.org/10.1155/2022/4435296.

Xu, L., Chen, W., Wang, S., Mohammed, B. S., & Lakshmana Kumar, R. (2022). Analysis on risk awareness model and economic growth of finance industry. *Annals of Operations Research*, 1–23. https://doi.org/10.1007/s10479-021-04516-z.

Yang, G. (2022). Research on financial credit evaluation and early warning system of Internet of things driven by computer-aided technology. *Computer-Aided Design*, *19*, 158–169. https://doi.org/10.14733/cadaps.2022.S6.158-169.

Yin, L.-L., Qin, Y.-W., Hou, Y., & Ren, Z.-J. (2022). A convolutional neural network-based model for supply chain financial risk early warning. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/7825597.

Yin, W., Kirkulak-Uludag, B., Zhu, D., & Zhou, Z. (2023). Stacking ensemble method for personal credit risk assessment in peer-to-peer lending. *Applied Soft Computing*, *142*. https://doi.org/10.1016/j.asoc.2023.110302. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85154065305&doi=10.1016%2fj.asoc.2023.110302&partnerID=40&md5=2882d057007a259a8972885b6013899e. Cited by: 2.

Yuan, K., Chi, G., Zhou, Y., & Yin, H. (2022). A novel two-stage hybrid default prediction model with k-means clustering and support vector domain description. *Research in International Business and Finance*, *59*, Article 101536. https://doi.org/10.1016/j.ribaf.2021.101536.

Yuping, Z., Jílková, P., Guanyu, C., & Weisl, D. (2020). New methods of customer segmentation and individual credit evaluation based on machine learning. In *New silk road: Business cooperation and prospective of economic development(NSRBCPED 2019)* (pp. 925–931). Atlantis Press.

Zand, S. K. (2020). Towards intelligent risk-based customer segmentation in banking. arXiv preprint, arXiv:2009.13929.

Zhang, W., Chen, R.-S., Chen, Y.-C., Lu, S.-Y., Xiong, N., & Chen, C.-M. (2019). An effective digital system for intelligent financial environments. *IEEE Access*, *7*, 155965–155976. https://doi.org/10.1109/ACCESS.2019.2943907.

Zhu, L., Li, M., & Metawa, N. (2021). Financial risk evaluation z-score model for intelligent iot-based enterprises. *Information Processing & Management*, *58*, Article 102692. https://doi.org/10.1016/j.ipm.2021.102692.