Mesopotamian journal of Cybersecurity
Vol.5, No.3, **pp.** 977–1041
DOI: https://doi.org/10.58496/MJCS/2025/055; ISSN: 2958-6542
https://mesopotamian.press/journals/index.php/cybersecurity

Systematic Review Article

Comprehensive study of Integrating Clustering and Adversarial Learning for Enhanced Recommender Systems: A Systematic Review of Hybrid Methodologies and Applications

M. E. Alqaysi[1], *, Murtadha M. Hamad[1], Ahmed Subhi Abdalkafor[1],

[1] *Computer Sciences Department, Computer Science and Information Technology College, Anbar University, Anbar, Iraq*

ARTICLEINFO

Article History

Received 13 April 2025
Revised 16 Jul 2025
Accepted 23 Aug 2025
Published 17 Sep 2025

Keywords

Robustness
Recommendation Systems
Clustering Methods
Adversarial Learning
Adversarial Attacks
Defence
Machine Learning

ABSTRACT

Recommender systems (RSs) have become critical elements in modern instances of information and decision-support systems, resulting in a transformation of user experiences through highly personalized suggestions for an undeniably vast range of items. Although RSs have become commonplace, they continue to evolve, and their challenges, including sparsity, cold-start, scalability, and vulnerability to adversarial challenges remain. The use of clustering methods has proven highly effective in resolving these issues through the discovery and exploitation of latent user behaviour patterns, segmenting user groups that contribute more towards personalized and adaptable RSs. Additionally, adversarial learning has become a growing focus of study as a proposed solution for shielding and defending RSs, processes, and data from manipulation and attacks, resulting in greater resistance and trustworthiness. This study presents a systematic literature review (SLR) that explores the intersection of RSs, clustering methods, and adversarial learning. This paper synthesizes a critique of the latest hybrid recommendations, detailing motivations, challenges, directions for future study, and practical recommendations drawn from the examined studies. An SLR search of four academic databases, ScienceDirect (SD), IEEE Xplore (IEEE), Scopus, and Web of Science (WoS), delivered an initial yield of 843 studies; after filtration, 51 studies remained. All the retained articles were examined and characterized in terms of dataset details, techniques, frameworks, and performance. A significant gap in research has emerged regarding the overreliance on datasets from commercial and entertainment domains, with a notable scarcity of studies addressing critical domains such as healthcare, finance, and other critical fields where diverse data sources should invoke robust, secure, and trustworthy RSs recommendations. Future research is needed to develop adversarially robust RSs for high-stakes applications requiring stringent accuracy and safety standards. This review provides a rich critical examination of the literature to embolden the ideas and theories associated with clustering methods and adversarial learning working together within RSs. It offers concrete opportunities and directions for carrying out future work in developing next-generation secure, adaptive recommendation frameworks. These findings corroborate a change in perspective on designing systems that seek to develop an RSs that can withstand adversarial threats and promote the development of safer, fairer, and more reliable decision-support systems in a variety of domains.

1. INTRODUCTION

Recommender systems (RSs) have become a vital part of modern digital ecosystems and have influenced the way in which users engage with content by providing targeted suggestions based on user preferences and behavioural patterns [1]. RSs on machine learning and data-driven approaches to analyse user interaction, purchase history, previous browsing behavior, and demographic features so that the platform can make relevant recommendations [2]. Companies and organizations across industries use RSs, including e-commerce, where they recommend products based on preferences (such as Amazon, and eBay); streaming services offer movie, music, or video content (such as Netflix, Spotify, and YouTube); healthcare, which provides medical treatment or medications based on patients' history; the education field, where services provide adaptive learning through recommending courses and other study materials (such as Coursera, and Udemy); and social networks, which find new friends and posts or professional connections (such as Facebook, and LinkedIn) [3]. RSs are divided into several approaches that primarily utilize collaborative filtering (CF) and content-based filtering (CBF), which are the two main strategies [4]. CF is based on the premise that users who have similar preferences will probably share similar interest’s preferences and lists items that users recommend, whereas item-based CF recommends items. CBF also

**Corresponding author. Email: mus22c1013@uoanbar.edu.iq*

uses a hypothesis of user preferences, but instead analyses item features and user preferences, recommending items that are similar to items the user has liked in the past. Both CF and CBF are functional methods for making recommendations but have serious drawbacks with respect to scalability and accuracy. Data sparsity is one of the obstacles to developing an effective RSs. The sparsity of data occurs when there are insufficient user-item interactions in the dataset, leaving too few for providing meaningful recommendations [5].

This issue of data sparsity is most prominent in large-scale systems, where most items have very few interactions, so the measurement of similarity may be flagged as unreliable. Closely related to data sparsity, the cold start problem occurs when a yet to be introduced user or item has no historical data. In this case, the RSs is unable to apply any recommendation methodology until sufficient user-item interactions are captured. Scalability issues also arise after RSs processes handle larger datasets; either the process cannot measure within the required time or the algorithms take too long to compute. In light of these challenges that have been mentioned. Here, an important question should be raised: **To what extent can the integration of clustering-based unsupervised learning techniques with supervised learning approaches for addressing address the limitations of traditional RSs, specifically while enhancing accuracy, personalization, and overall system robustness?**

A key component of these systems is machine learning, an impressive collection of models and algorithms that enables computers to learn patterns in the data and make intelligent decisions. Supervised learning is an important technique for predicting a user's preferred options and provides personalized recommendations [6], [7], [8].

Supervised learning employs labelled data (which is an input feature and an output label that is associated with the result), where an example is a user's explicit recommendation for a movie or their decision to purchase an item [9], [10], [11]. There are many supervised algorithms that have been successfully applied in user behaviour modelling: decision tree (DT), random forest (RF), support vector machine (SVM), and k-nearest neighbors (KNN) [12], [13].

DT use a strategy by partitioning the feature space into subregions where each subregion is defined by a feature's value [14]. DT are tree-shaped structures that are easily interpretable and allow complex non-linear relationships to be captured. This property of interpretability gives the system designer insights related to providing recommendations on which user- or item-based attributes are most influential in constructing recommendations. RF builds on the idea behind a DT by building an ensemble of multiple trees trained on a bootstrap sample of the data while choosing a random subset of features for each tree [15].

RF method offer advantages against variance and improves generalization. SVM find a hyperplane that maximizes the separation between different classes in a multidimensional feature space and are very effective if the task is a binary or multiclass classification task (e.g., predicting if a user will like or dislike an item). KNN operates on the basis of similarity, such that KNN classifications occur based on the majority label of the k closest neighbors in the feature space [16], [17].

KNN algorithms are easy to understand and very suitable for CF recommendations because an explicit user rating of a similar user or characteristics of a similar item is usually available. Supervised learning is advantageous for precision in repeated predictions with feedback of a known value, but supervised learning methods work equally well in RSs and learning by discovering hidden structures and hidden patterns from data that are not apparent [18].To identify these hidden structures from data, clustering approaches such as K-means, DBSCAN (density-based spatial clustering of applications with noise), deep embedded clustering (DEC) and variational deep embedding (VaDE) are helpful for this purpose [19].

K-means is one of the more straightforward machine learning algorithms, and is the most recognized and commonly used clustering form, k-means clustering divides users or items into k clusters while trying to minimize the variance within the clusters, in essence, grouping users with similar preferences or items with similar features for a recommendation system where a group-based identified user recommendation may provide more relevance to the user in a focused manner to personalize the recommendation system to that group [20].

DBSCAN, in contrast, defines a cluster as a collection of dense regions of data space separated by lower-density regions. DBSCAN is capable of identifying clusters of arbitrary shapes, as well as outliers; therefore, it is useful for segmenting users with niche or unusual behaviours that may not be detected via a standard clustering approach. DEC combines clustering and deep learning and learns feature representations important for the task of clustering and the cluster assignments in an autoencoder [21].

The DEC method can learn a compact and informative embedding of user/item data, as well as develop equivalent, although complex, relationships that might not be available from traditional clustering methods. This is significant in recommendation type environments where user behaviour could be based on an amalgamation of factors and interactions suggesting how DEC can learn a multidimensional representation of user/item data. VaDE extends this concept by adding a vibrational autoencoder to the clustering learning framework while also relying on a Gaussian mixture model. In this way, the system can reflect the complex but overlapping interests that a user may have. An additional benefit of a hybrid of supervised and unsupervised learning provides an additional method for developing an application recommendation, where it relies not only on accurate predictions, but also on uncovering hidden patterns in user use. While supervised models depend on historical feedback, they face the difficulties of new users and sparse data. Unsupervised clustering helps segment users based on implicit behaviours, providing a structural foundation that enhances supervised predictions even with limited data. This combined strategy allows systems to adapt dynamically as user preferences evolve, leading to more accurate, personalized, and diverse recommendations.

Ultimately, this synergy enables the RSs to better understand and predict user preferences, supporting robust and user-cantered methods. The combination of traditional and deep clustering methods will also help reduce the cold start problem, improve computational efficiency, reduce computational load, and improve personalization, leading to improved accuracy and increased user experience. In addition to the previous consideration of how to utilize RSs, one other area that has considerable potential to move the recommender system field forward is adversarial learning. This is especially the case for flexibility and robustness, which most recommendation systems suffer from. In that supposition comes the following consideration: **What is the role of adversarial learning?**

Adversarial learning has emerged as a powerful and comprehensive framework in artificial intelligence (AI), serving as a "big umbrella" under which key concepts such as adversarial attacks and adversarial training are integrated to strengthen system security and robustness. In the context of an RSs, which plays a critical role in guiding user choices across e-commerce, media, and social platforms, adversarial learning provides essential mechanisms to address vulnerabilities and enhance reliability [22].

Adversarial inputs are examples where the input feature representation may be different, intentionally modified, or altered, which includes some characteristics that mislead the models' true prediction or ultimate learning [23]. These features are typically modified in subtle ways that humans cannot detect but can have large implications for model performance [24].

Adversarial attacks involve deliberately crafted perturbations designed to deceive recommender models, causing them to generate irrelevant or misleading suggestions, and thus serve as a diagnostic tool to uncover hidden weaknesses in user–item interaction patterns [25], [26].

Common techniques are used to generate these perturbations and include fast gradient sign method (FGSM) attacks, projected gradient descent (PGD), and Carlini & Wagner (C&W) attacks, which each slightly alter the input signals but still induce model failures [27].

As an example of how to defend against attacks, adversarial training incorporates adversarial examples into model training, allowing the recommender system to learn robust representations of the input data and build resistance against manipulations. In this context, models can employ PGD-based adversarial training and TRadeoff-inspired Adversarial DEfense via Surrogate-loss minimization (TRADES), which uses an appropriate mix between natural accuracy and robustness, enabling models to maintain performance under adversarial conditions and to adapt simultaneously to changing user behaviors [22], [28].

Aside from the attack and defense dichotomy, adversarial learning can extend further into other wider spheres, such as generative adversarial networks (GANs), which can be used to generate synthetic user behavior data to improve recommendation diversity and mitigate data sparsity issues [29].

Similarly, adversarial domain adaptation techniques are able to help recommendation models generalize across different user groups or markets by jointly learning invariant representations of features. By integrating these adversarial learning techniques, the RSs not only becomes more secure and robust but also achieves higher generalizability, fairness, and adaptability. This holistic approach encourages researchers to explore its transformative impact on creating more trustworthy, user-centered, and resilient recommendation solutions that can withstand real-world adversarial scenarios [30]. The application of adversarial learning increases models' resistance and robustness to those manipulations, which can improve their ability to deal with noisy, sparse, and manipulated data. Now, a critical move in research is to ask: **How can adversarial learning be leveraged to enhance the robustness of clustering-based RSs against adversarial attacks, ensuring both accuracy and security in personalized recommendations?**

By using adversarial learning techniques, models can therefore train more effectively to identify and reduce adversarial mechanism threats and adapt to them, with better robustness and generalized performance capabilities [31].

In practice, adversarial training integrates these crafted adversarial examples in the learning process of a model, optimizing the model with a min–max objective to minimize the prediction error of the model while using the worst-case perturbations. This would better promote more stable cluster structures and is less likely to lead to data poisoning attacks, in which malicious data points are included to intentionally bias the recommended output of the recommendation system. Furthermore, the incorporation of broader adversarial learning strategies, such as GANs, can provide synthetic adversarial user behavior data that simulate diverse, realistic, and unexpected interactions.

This augmentation improves both the representativeness of training data and the system's ability to generalize to new user behaviors or attack strategies. Combined with the characteristics of RSs and clustering algorithms, this technique has the potential to solve vulnerabilities and improve systemic resilience to manipulative attacks and data inconsistencies [32].

In addition, adversarial domain adaptation methods can help clustering-based recommenders maintain consistent performance across different user groups or markets by learning domain-invariant features, thereby reducing sensitivity to adversarial shifts in user data distributions. While these approaches improve safety and robustness, they also introduce challenges such as higher computational costs and tuning of perturbation strengths. Nonetheless, integrating adversarial learning into clustering-based RSs represents a promising direction for building a valuable avenue to explore, especially in scenarios requiring secure operation during an adversarial shift that can be described as being extremely dynamic.

2. METHODOLOGY

In this study, a systematic literature review (SLR), which is one of the most recognized and systematic approaches to synthesizing prior literature in a specific research area, was conducted [33], [34], [35], [36] The SLR approach was selected because it provides comprehensive and objective checks of relevant literature when identifying, selecting, and analysing studies based on predetermined criteria. SLR is different from traditional literature reviews since they are conducted in a manner that minimizes bias and increases the reproducibility of study findings when reputable databases are used. An SLR provides a framework for organizing research, as opposed to unorganized research, so that it provides a comprehensive examination of a specific research area.

To ensure broad and reliable coverage of literature, multiple databases were selected, as relying on a single database may introduce selection bias or omit critical studies. On the basis of their relevance and credibility in the domain of modern research, four major academic databases were chosen:

1. Web of Science (WoS): a world leader in research databases and includes many of the highest impact publications across all fields. WoS is essential for reliable and solid study evidence, as it includes only peer-reviewed journals.
2. ScienceDirect (SD) includes a large and varied scientific journal database covering disciplines, including but not limited to, inclusive but not limited to, medicine, science, and technology, to include and support interdisciplinary research and contribute information to these domains.
3. IEEE Xplore (IEEE): As the name implies, IEEE is primarily oriented toward engineering and technology-related research, and along these lines, IEEE is well known for its extensive collection of high-quality publications, particularly in the areas of computer science, AI, and machine learning.
4. Scopus: As one of the largest abstract and citation databases, Scopus has impressive coverage of scientific literature and quite an extensive amount of nonscientific and scientific literature, allowing for a broad analysis of applicable studies.

2.1 Search Strategy

For this investigation, four databases were thoroughly searched for English-language scholarly literature. The search ranged from January 2019 to January 2025. A keyword-based query was formulated to systematically retrieve relevant studies, as demonstrated in Figure 1. This structured search strategy ensures the inclusion of relevant literature while filtering out studies that do not align with the research focus. The "OR" operator was used to connect "clustering deep", "clustering", "adversarial attack", "adversarial learning", "recommender system", "recommendation system", "machine learning", "deep learning," and the "AND" operator linked these phrases together, as shown at the top of Figure 1. This search approach was used to find relevant scholarly literature.

2.2 Inclusion and Exclusion Criteria

To maintain the quality and relevance of the reviewed literature, predefined inclusion and exclusion criteria were applied during the selection process.

- Inclusion Criteria

The selected studies had to meet the following conditions:

1. Published in peer-reviewed English-language journals or conference proceedings.
2. This is directly related to more than one of the following research areas: RSs, clustering approaches, adversarial learning, and AI-driven methodologies for enhancing algorithmic performance.
3. Focus on integrating or advancing the above concepts through novel techniques, frameworks, or applications.

- Exclusion Criteria

Studies were excluded if they met any of the following conditions:

1. Did not explicitly address the intersection of RSs, clustering, and adversarial learning.
2. The concepts were used only as secondary considerations rather than as core research themes.
3. Were purely medical studies.
4. Studies that use only a traditional recommender system.
5. Studies that use clustering or adversarial learning only.

By implementing these rigorous selection criteria, the study ensures the inclusion of only the most relevant and high-quality research, thereby strengthening the validity of the findings and contributing to a well-founded analysis of the field.

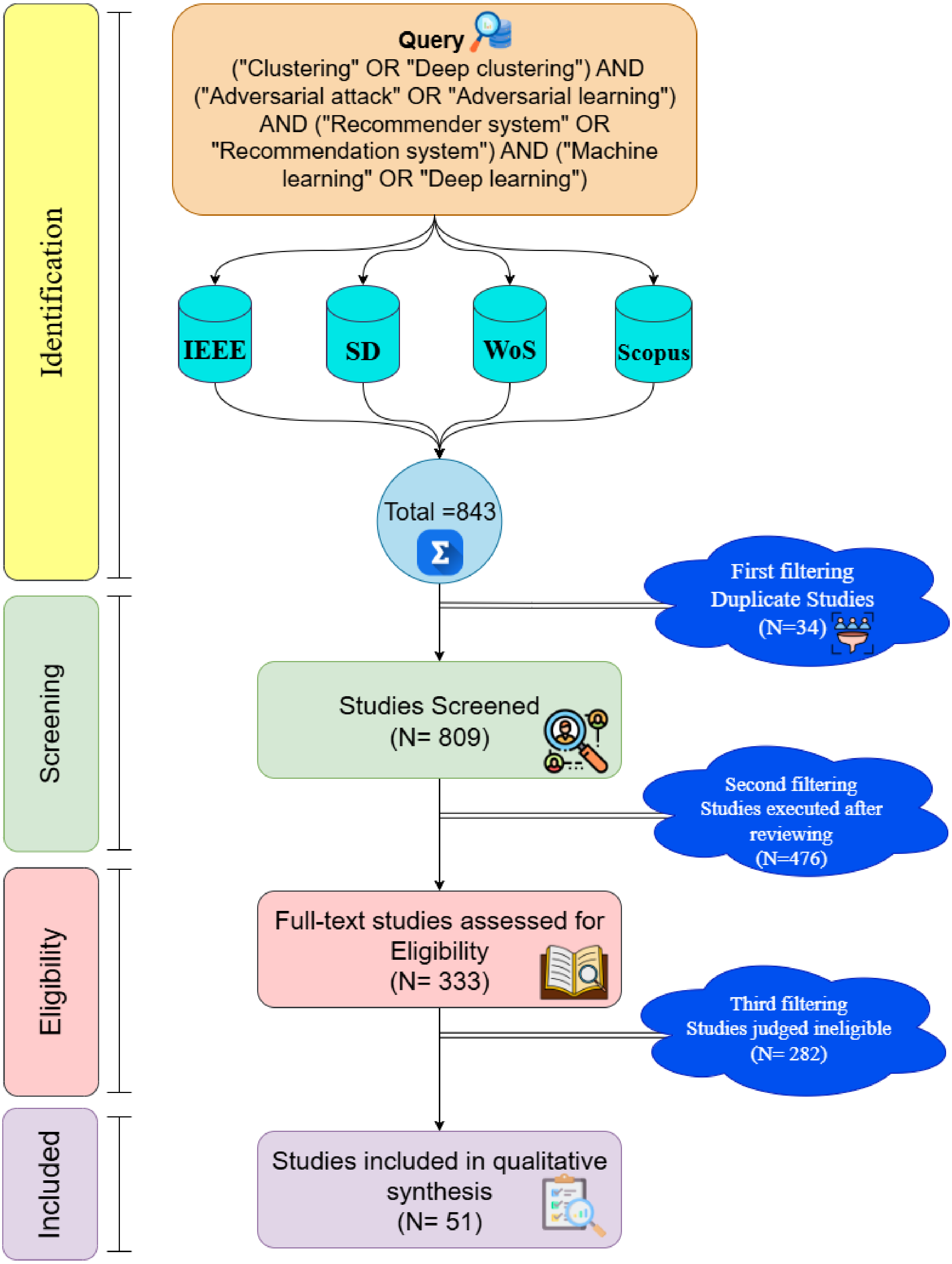
Identification
Screening
Eligibility
Included
Query
("Clustering" OR "Deep clustering") AND ("Adversarial attack" OR "Adversarial learning") AND ("Recommender system" OR "Recommendation system") AND ("Machine learning" OR "Deep learning")
IEEE
SD
WoS
Scopus
Total =843
Σ
First filtering
Duplicate Studies
(N=34)
Studies Screened
(N= 809)
Second filtering
Studies executed after reviewing
(N=476)
Full-text studies assessed for Eligibility
(N= 333)
Third filtering
Studies judged ineligible
(N= 282)
Studies included in qualitative synthesis
(N= 51)

Fig. 1 An outline of the approach used to identify, select, and include relevant

2.3 Study Selection Process

The study selection process followed a systematic and rigorous methodology, as outlined in Figure 1. The initial step involved the removal of duplicate studies via the Mendeley reference management application to ensure a refined dataset. The titles and abstracts of the retrieved studies were subsequently screened to exclude irrelevant works that did not align with the research objectives. The full-text data were then reviewed in accordance with established inclusion criteria to decide whether they were appropriate.

A total of 843 research articles from the four pooled databases. After 34 duplicates were removed, 809 papers remained eligible. The titles and abstracts of the 809 articles were then screened, and 476 were removed because they were clearly not relevant to this study. A full-text examination of the 333 papers remaining led to the removal of 282 studies, leaving 51 studies for comprehensive evaluation and in-depth analysis.

2.3 Data Extraction and Classification

Through a systematic review process, relevant information was carefully extracted from the qualitative analysis of the included papers, with a structured subgroup review enabling the analysis, which focused on important variables, including the type of AI technology, algorithm types, performance metrics used to assess AI model performance, datasets available, and main contributions. A comprehensive full-text review was performed, and each selected article was selected according to the aims of the research. The classification of the studies reviewed was also accomplished iteratively, forming the basis of the taxonomy defined in this study. This study also synthesized a range of literature related to research on RSs, clustering and adversarial learning while providing dimensions related to aims, motivations and challenges associated with researchers working in this domain. The review also summarized any limitations from previous research, and this study provided recommendations for future research, which are described in later sections. This systematic approach synthesized the literature and identified gaps and opportunities for future work in this area.

3. COMPREHENSIVE SCIENCE MAPPING ANALYSIS

In this section, as presented four in-depth analysis procedures that we developed from studies taken from four databases. The study represents a detailed analysis of the objectivity of the research papers by considering various factors encompassing the research methodologies, quality and extent of the studies, impacts (and citation counts) of these articles and linking the articles and results to a wider research base in various disciplines. Through an analysis that considers these factors, we analysed the studies, the patterns and trends in study designs, and emerging topics within a research topic area, as well as the methodological rigor implied by the study articles, which contributed to knowledge in their respective areas of research. In the following subsections, we discuss in detail the analysis of these studies and the benefits they offer on the basis of the most relevant source, Word Cloud Analysis, Countries Scientific Production and Collaboration Map, and Co-occurrence Network.

3.1 Most relevant sources

The provided data illustrated in Figure 2 show a range of academic journals and conferences in different proportions during the publishing process. The IEEE Transactions on Knowledge and Data Engineering produced four articles that are at the top of the list, followed by journals such as IEEE Access, Information Fusion, and Knowledge-Based Systems, each of which had three articles included in the data extraction. This indicates their academic rigor in publishing three articles on the basis of the value of a criterion for a significant emphasis in fields related to AI and data engineering. Expert systems with applications, neural computing, and applications, while contributing two publications each, are likely to have a strong but niche reputation with lower publication rates. Conferences such as AAAI 2020 offer prestigious platforms for researchers, despite featuring only one article in the dataset. This underscores the importance of conference papers in top-tier research. Journals with a single publication, such as Computers and Security and Pattern Recognition, also maintain relevance in their specialized fields. However, overall, the data show that the publication volume of articles does reflect active research, and journals and conferences valued in publications (for example, IEEE and Elsevier) have significant value in well-published, distributed, and impactful research, especially in terms of AI, security, and data analytics.

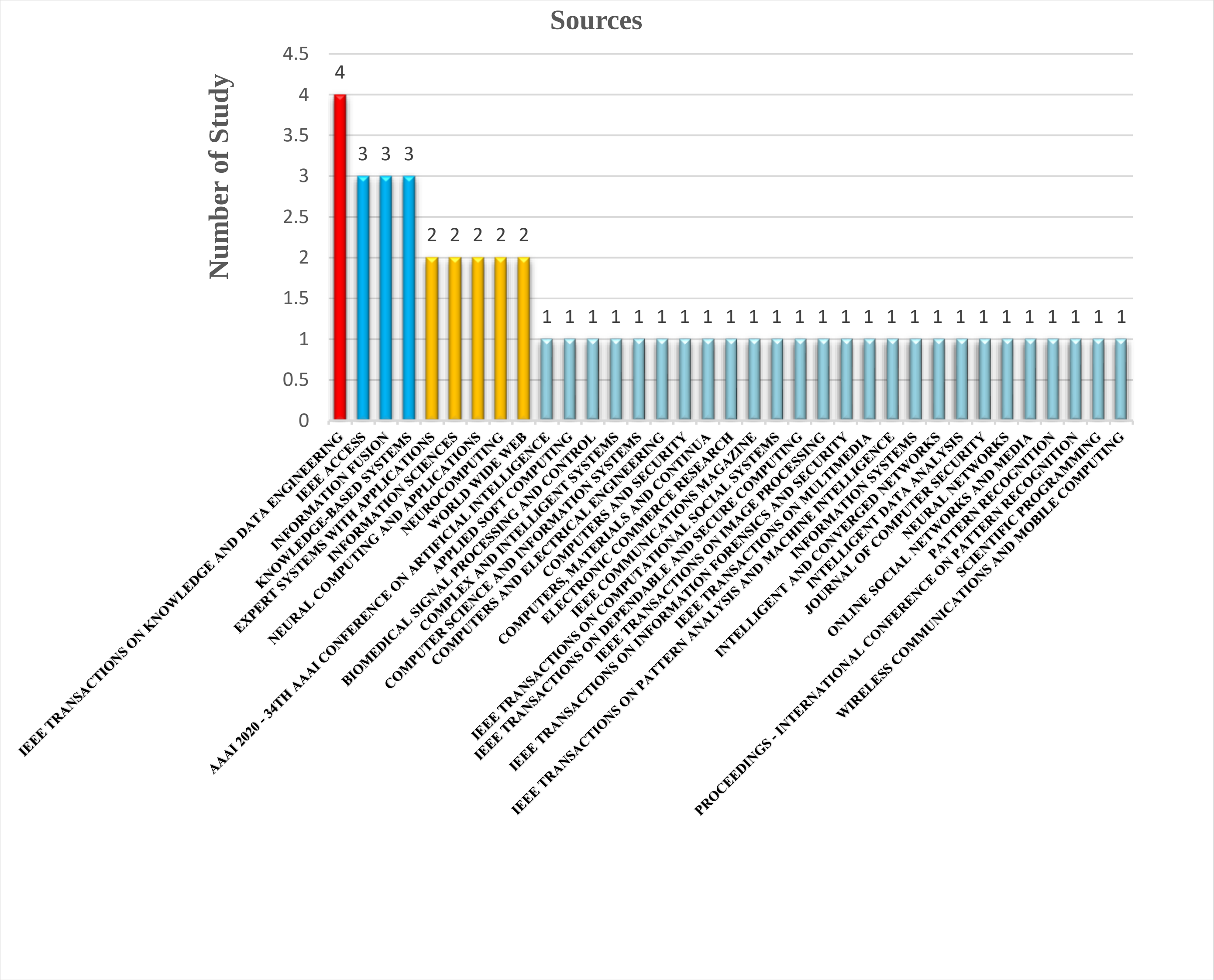

Fig. 2 Most relevant sources

3.2 Word Cloud Analysis

Figure 3 illustrates a visual representation of the key themes and data used in the studies analysed. It is a good format for summarizing and highlighting the main topics associated with a certain research area. By using word clouds to visually represent the most common terminology through font size and central positioning, specialists can clearly understand and easily identify important concepts. In this study, the word cloud was generated on the most frequently appearing keywords across the reviewed literature, offering insights into the primary research focus areas. The prominence of a keyword within the word cloud is directly proportional to its frequency in the dataset, with more frequently occurring terms appearing larger and more centrally positioned. This visualization aids in identifying dominant trends, recurring themes, and critical areas of interest in the selected research corpus. Figure 3 shows a word cloud constructed from the 143 most commonly used keywords in the analysed studies, providing a comprehensive overview of the primary subjects addressed in the literature.

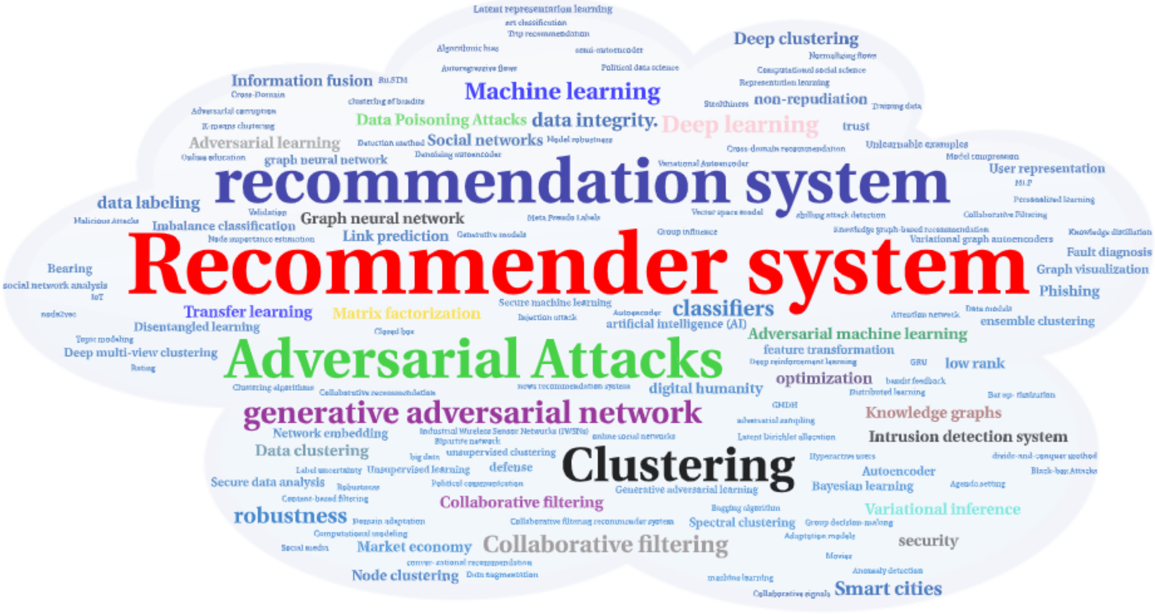

Fig. 3 Word Cloud Analysis

3.3 Countries Scientific Production and Collaboration Map

Figure 4 shows that scientific research plays a crucial role in advancing knowledge and driving technological innovations in this study's collection, with China leading global scientific production, followed by the USA and India in that direction. A dark blue color represents a high level of research output, whereas light and very light blue colors indicate lower levels of scientific productivity. China plays a dominant role in advancing RSs, adversarial learning, and clustering, producing a significantly greater volume of research than other nations do. Both the USA and India have also made significant contributions to this area and continue to be influential players in AI-led analytics. However, other countries, such as Italy, Germany, and Morocco did not produce much in this work, indicating a lower research presence. Even though Germany and Italy possess a well-established technological infrastructure, their scores on AI-based research were low, indicating little research activity. The pace of change in countries like China and the USA has made it exceedingly challenging for countries like Germany and Italy to keep their research base up to speed. Clearly, a lack of research contributions in countries such as Germany and Italy indicates that AI-led innovation and analytics is not a research priority or investment area, which could risk their competitive advantage in future global technology. Moreover, collaboration in scientific research on an international scale is critical for enhancing research and innovation globally. The value of collaboration and knowledge transfer is used to assess the quality of scientific knowledge and research projects. Therefore, it is not surprising to see the rankings of China and the USA as collaborative countries to one another, where the strength of their collaborative support is represented as the gray line connecting the two great research networks and flagship contributors to the global research output. The thick gray line indicates an increase in research collaboration. The thicker the line is, the greater the intensity of research collaboration between the countries, and vice versa. China has the highest value of international collaboration by a fair amount, with noteworthy participants from Australia (4), Hong Kong (4), the USA (5), and the UK (2). This indicates the potential growth of China's influence on global research trends in areas such as AI and data analytics. The USA has also worked with Hong Kong (2), where notable partnerships exist with India (1), Pakistan (1), and Poland (1), showing that the USA is still a significant factor in the exchange of knowledge worldwide. Other European countries, such as Austria, Ireland, and Poland, are actively involved in the limbs of large-scale international investigations but show less international collaboration.

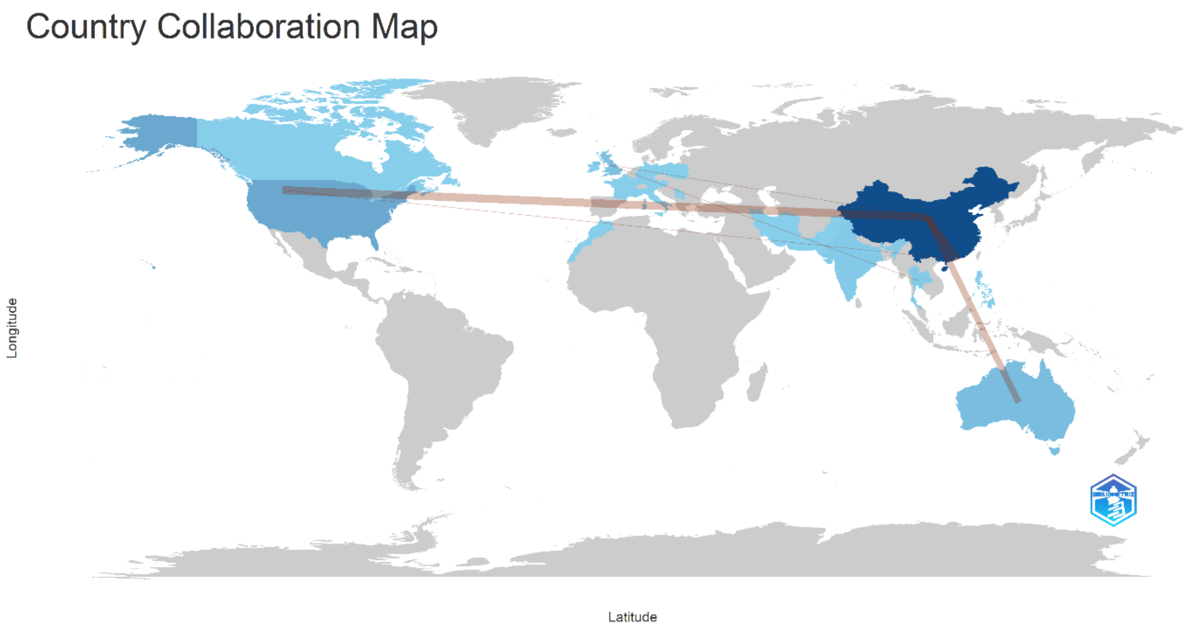

Fig. 4 Provincial Production and Collaboration Map of Countries

3.4 Co-occurrence Network

The co-occurrence network captured in Figure 5 illustrates RSs as an essential area of research, connected to key methods such as collaborative filtering, clustering methods, and knowledge or semantic graphs as techniques for producing personalized experiences that rely on behavioral analyses (i.e., user profiles, clustering) and contextual semantic reasoning (e.g., graph-based embeddings). These systems rely on machine learning methods, including deep learning and graph neural networks, to refine recommendation processes. However, they also encounter challenges such as adversarial risks, including poisoning attacks aimed at corrupting training data, as well as ethical concerns, such as the potential for bias amplification. Interdisciplinary connections are evident in the use of knowledge graphs, which improve recommendation accuracy by organizing relational data, and in the application of human-centered technologies, such as e-learning, where insights from behavioral research guide the development of adaptive and personalized algorithms.

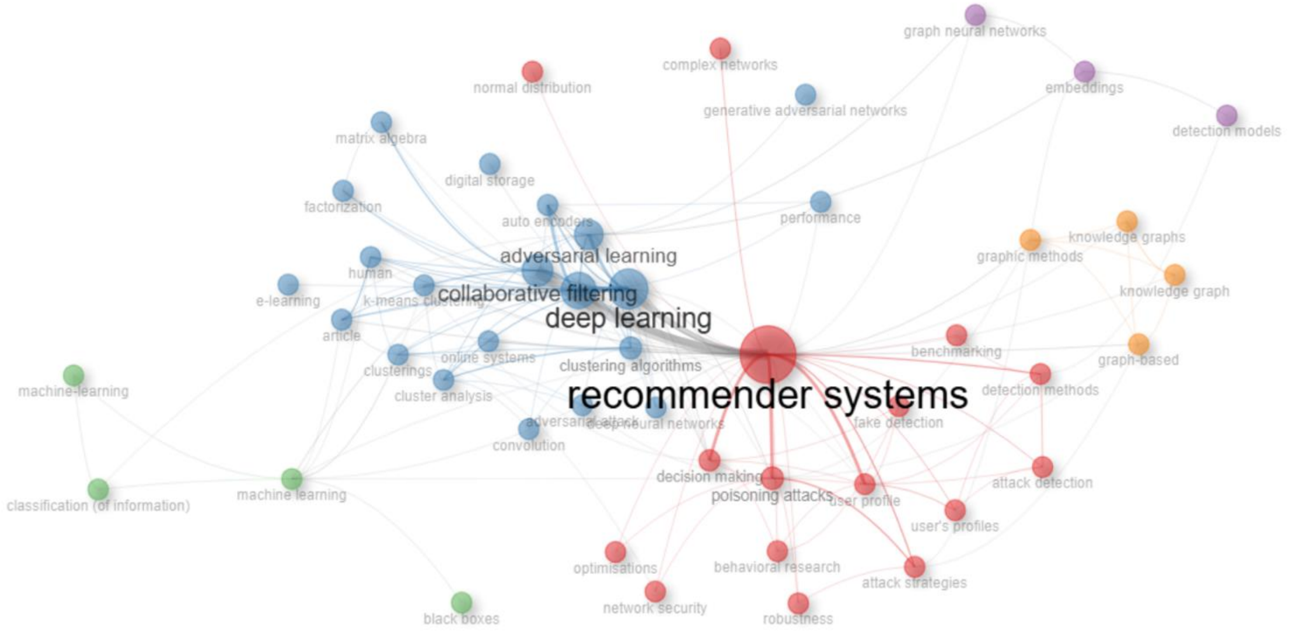

Fig. 5 Co-occurrence Network

4. TAXONOMY RESULTS

This taxonomy is structured around four principal sections, each comprising subsections that delineate the subclassifications derived from an extensive analysis of studies collected from academic databases. The objective is to distil the fundamental and influential elements that shape the foundation of this research, uncovering key insights and conceptual frameworks that define this evolving field. The total number of extracted studies amounts to 51, forming the empirical basis of this classification. As illustrated in Figure 6, this structured methodology ensures a systematic and evidence-driven analysis, grounded in objective scholarly inquiry. At its core, this taxonomy is built upon the intersection of clustering, recommendation systems, and adversarial learning, reflecting the intricate interdependencies that characterize these domains.

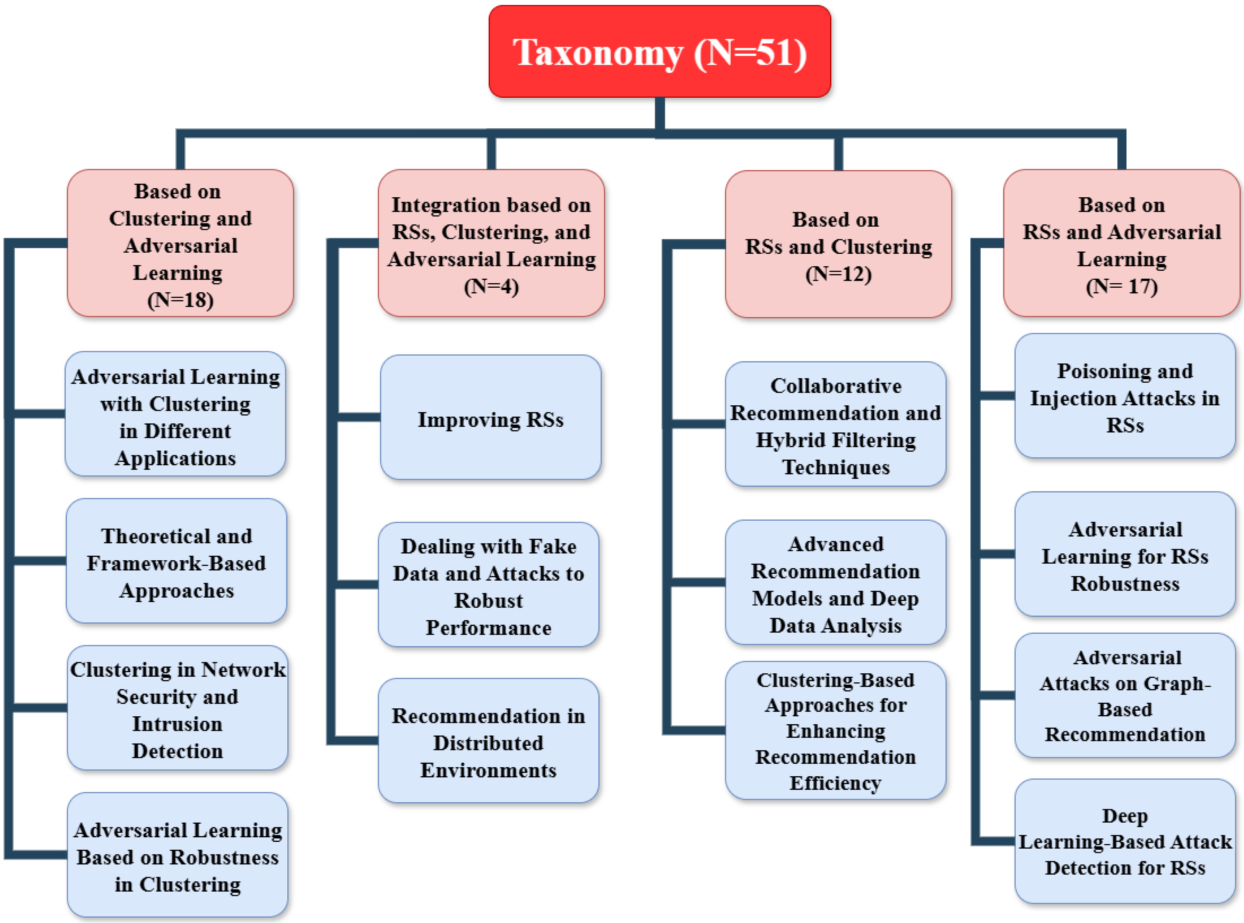
Taxonomy (N=51)
Based on Clustering and Adversarial Learning (N=18)
Adversarial Learning with Clustering in Different Applications
Theoretical and Framework-Based Approaches
Clustering in Network Security and Intrusion Detection
Adversarial Learning Based on Robustness in Clustering
Integration based on RSs, Clustering, and Adversarial Learning (N=4)
Improving RSs
Dealing with Fake Data and Attacks to Robust Performance
Recommendation in Distributed Environments
Based on RSs and Clustering (N=12)
Collaborative Recommendation and Hybrid Filtering Techniques
Advanced Recommendation Models and Deep Data Analysis
Clustering-Based Approaches for Enhancing Recommendation Efficiency
Based on RSs and Adversarial Learning (N= 17)
Poisoning and Injection Attacks in RSs
Adversarial Learning for RSs Robustness
Adversarial Attacks on Graph-Based Recommendation
Deep Learning-Based Attack Detection for RSs

Fig. 6. Taxonomy of the use of the intersection among clustering, recommendation systems, and adversarial learning

4.1 Based on RSs and Clustering

This section presents twelve research papers (12 in total) that explore the intersection of RSs and clustering techniques. This intersection plays a pivotal role in improving the accuracy, scalability, and personalization of recommendations in various domains. Clustering techniques help organize users or items based on similarities in their properties into groups with different characteristics or behaviors. Together, these studies cover the effectiveness of combining clustering with recommendation models in improving system performance and recommendation quality.

4.1.1 Clustering-Based Approaches for Enhancing the Recommendation Efficiency

The studies included in this subsection consist of five research papers (N=5) that specifically focus on the intersection of clustering techniques and RSs, which have made significant improvements in the same areas of personalization, efficiency, and scalability in a plethora of applications. The intersection of clustering techniques and recommendation systems has resulted in transformational improvements in personalization, efficiency, and scalability in various applications. By leveraging clustering techniques, modern RSs can improve the organization of user–item interactions, ultimately improve computational complexity, and address certain inherent challenges, such as cold-start challenges, data sparsity, and long-tail distributions. The field of news recommendation can serve as an example of this, where a newly published recommendation model integrates graph neural networks with bat optimization algorithms to improve news articles' clustering abilities by acquiring semantic information; in turn, this approach addresses the issue of variety in the popularity of news articles in a news recommender system [37]. Furthermore, the use of attention mechanisms generalizes and streamlines the process of optimizing the content of interest, vectorizing users' news preferences, and providing better resolution recommendations. In types of applications beyond content-specific purposes, clustering techniques/frameworks based on network embeddings (e.g., N2VSCDNNR) create side information in terms of item categories to mitigate the potential problem of sparsity [38]. As automatic cluster number determination is a method based on a normal distribution, relying on normal distribution and confidence intervals, it enables a dynamic clustering strategy, reducing the time complexity of real-time recommendations while improving accuracy. In another direction of the Internet of Things (IoT), clustering-based recommendation systems harness techniques such as k-means, fuzzy c-means, single-linkage, and self-organizing maps to address the challenges of scalability, sparsity, and diversity [39]. These approaches significantly outperform traditional CF models in IoT-driven environments, enhancing the adaptability of recommendation frameworks. On the other hand, when addressing user cold-start problems, particularly in online movie networks, another clustering-based methodology integrates content-based filtering, collaborative filtering, and data mining techniques [40]. By embedding a clustering mechanism based on similarity, this approach provides more relevant movie recommendations to new users while improving the quality of their first recommendations. Finally, the use of clustering-driven deep learning in online learning platforms denotes the effective use of clustering techniques to assist in processing large amounts of varied educational data and building accurate recommendations [41]. These models offer more personalized, scalable, and adaptive course recommendations by dynamically clustering or segmenting learner profiles and course content.

4.1.2 Advanced Recommendation Models and Deep Data Analysis

As examined two research papers (N = 2) in the preparation of this subsection. RSs have become more prevalent, and recommender models have evolved to meaningfully extend their understanding from personalization and user experience. Not only does the Group influence mechanisms and cross domain recommender strategies provide flexibly structured framework in improvement to recommendation quality, it is aware of context. The group influence-aware autoencoder (GI-AAE) model also offers a new improved means of creating top-N recommendations, where the latent feature representation is strengthened using group interactions [42]. Not only is information fusion becoming better and group-based decision-making added, the model is now capable of producing situationally relevant recommendations. This model also takes quality to the next dimension extracting latent features is now clearer, and recommendations in sparse systems are now accurate [19]. When group-based modelling techniques and cross-domain context are combined, contextual factors produce a more personalized and user-centric recommendation experience. In this way, improved recommender models enhance system intelligence, scalability, and flexibility.

4.1.3 Collaborative Recommendation and Hybrid Filtering Techniques

This subsection uses data from five selected research papers (N =5). As recommendation systems have advanced, there has been a shift in the mixing of CF with hybrid techniques to increase the accuracy, scaling, and personalization of recommendations. This includes advancements in newer domains of probabilistic models, deep neural networks, and knowledge graph-based approaches that allow for a systematic refinement of interaction factors for users and items to enhance prediction performance. One example is a Collaborative Autoregressive Flows model that uses a Bayesian inference framework for probabilistic recommendations [5]. Autoregressive flows provide the flexibility to improve how posterior approximation is modelled for more accurate and interpretable recommendations; a similar refinement process is the hybrid collaborative recommendation via a dual autoencoder. It is similar to traditional CF but incorporates matrix factorization in its autoencoder training to improve the quality of representations in its hidden features [1]. This hybrid technique also makes CF less computationally expensive and stigmatizes the cold-start problem by adding user and item attributes. Further to hybrid techniques, a study incorporates content and CF in order to combine techniques [4]. This type of hybrid recommendation algorithm utilizes models focused on the exploitation of recommendations, ensuring the reduction of data sparsity and runtime concerns while increasing the number of effective recommendations. On a related subject, movie recommendation systems, which are subject to CF models, focus on evaluating the content in the CF models. Hence, the recommendations are more thoughtfully matched to user preferences, and effort is minimized in terms of content reduction [43]. In addition, the CSEKG knowledge graph-aware model implements a collaborative signal injection mechanism, which includes node importance estimation and item clustering based on CF [44]. This approach provides additional user–item relations and enhances the quality of recommendations via the structured application of knowledge. By embedding probabilistic models, deep learning architectures, hybrid filtering methods, and knowledge graphs, these approaches support the next generation of RSs. The implementation of these approaches in combination results in each contributing to higher accuracy, scalability, and user experience, which are the requirements of recommender system development. As such, these approaches can be classified under RSs in machine learning.

4.2 Based on Clustering and Adversarial Learning

This SLR reviewed the area of clustering in conjunction with adversarial learning using eighteen articles (N=18). Clustering is an important unsupervised learning approach. Clustering is widely used in pattern recognition, anomaly detection, segmentation of data, and so on. Clustering even though an unsupervised learning approach is prone to adversarial attacks that impact the robustness and security of clustering. Adversarial learning is a technique that arises from the inconvenience of defining adversarial attacks to improve resilience to clustering and examine the flaws of clustering. This literature review categorized research into application-derived clusters such as network security, trust management, and so on. We explored both theoretical advances and application advances. The model is created as follows. In the subsequent subsections, we provide a detailed description of the potential taxonomy created from the studies analysed.

4.2.1 Adversarial Learning Based on Robustness in Clustering

The incorporation of adversarial learning into clustering models specifically emphasizes the robustness of clustering models against adversarial attacks. The research works included in this subsection consisted of seven research works (N=7). Clustering methods (e.g., K-means) are crucial to machine learning, since these methods entail grouping ‘n’ number of data points according to their similarities. However, these methods are susceptible to adversarial perturbations, which can have a crippling effect. Research that focuses on adversarial learning aims to address the susceptibility of clustering methods to adversarial attacks by providing information about the impact of adversarial learning on clustering and proposing solutions to improve the robustness of clustering algorithms. The first research work addresses black-box adversarial attacks on clustering algorithms, specifically on linearly separable clusters [45]. This work outlines a scenario in which perturbing a single sample close to a decision boundary could affect many close, unperturbed samples by changing their cluster assignment into the perturbed sample's cluster, i.e., spill-over adversarial samples. Research has shown that it is possible to attack clustering methods adversarially, and it does not require knowledge of the actual metric used to cluster the samples, revealing the hidden dangers of clustering methods. The second study investigated the robustness of clustering algorithms to adversarial noise. A black-box adversarial attack is proposed to solve this research problem as a constrained minimization program [46]. This approach evaluates the durability of clustering algorithms by creating adversarial samples tailored to what the attacker can do. The research also analyses how vulnerable different clustering algorithms are to adversarial samples as they have been created, benchmarks state-of-the-art approaches, and highlights the necessity of improved algorithms in these adversarial contexts. The third study focuses on adversarial attacks, specifically decision-time and data poisoning attacks against clustering models. The research looks at clean-label poisoning attacks, performing data poisoning on our training data, where the training data originally came in a clean label part of the approach by adding small perturbations to every training data point to illustrate how adversarial attacks can reduce performance within clustering models [28]. The study attempts to emphasize how clustering approaches are susceptible to such adversarial attacks while also emphasizing the need to improve our clustering models in order to defend against this type of attack. To address these adversarial threats, another study proposed a robust clustering method using an attention mechanism and graph convolutional networks (GCNs). This facilitates a more effective blend of nodal features and topological structures to improve clustering performance. Additionally, we propose a graph purification method that has a defense mechanism against adversarial attacks on graph data, thereby strengthening the overall robust nature of clustering approaches [47]. However, some studies provide evidence of how GANs perform in clustering models, as they relate to identifying consistency in information, representing identification with high-dimensional data, and recognizing the appropriate number of clusters with deep learning clustering models. One of the main studies that provides a significant impact to the field of clustering via adversarial learning is alternating generative adversarial representation learning (AGARL), a new multiview clustering framework based on an alternating generative adversarial strategy [48]. The AGARL framework maintains and enhances the ability of clustering methods to achieve a high degree of clustering performance by identifying and synthesizing clusters across views while leveraging consistent information across multiple views. AGARL outperforms shallow and deep multiview clustering in empirical studies conducted on publicly available datasets, demonstrating the robustness of the framework in clustering heterogeneous representations of data into a single view to make sense of the data representations. The other research consists of the eClusterGAN model, which is intended to improve clustering with latent space to recognize more efficient methods to analyse datasets, similar to other processes in complex market economies [49]. The final study within this subsection addresses a fundamental issue in clustering automated cluster number determination within the context of spectral clustering. By integrating spectral clustering with GANs and low-rank models within a Bayesian framework, this research introduces an adversarial-learning-based deep clustering method. A key innovation in this approach is the incorporation of a hidden space structure preservation term, which enhances the generative process and ensures more precise and scalable spectral clustering outcomes [50].

4.2.2 Clustering in Network Security and Intrusion Detection

This section contains three studies (N=3) that investigate clustering approaches for improving network security, namely, network intrusion detection systems (NIDSs) and phishing identification mechanisms. The studies in this section examine how clustering approaches can be used to further leverage the resilience and accuracy of network security mechanisms in adversarial situations. The first study highlights the class imbalance issues found in NIDS due to the inability to distinguish legitimate traffic, direct attacks or obfuscated intrusions and discusses a multiple-clustering-based under-sampling framework that improves classification accuracy based on selecting representative centroids from clusters [51]. In contrast to feature vectors, the machine learning model is not only accurate but also recognizes unseen malicious instances, thereby increasing the ability of the NIDS to adapt to recognize future unknown intrusion scenarios. The second study contributes to the literature on phishing detection mechanisms, demonstrating their robustness towards adversarial learning attacks. The study shows how adversarial attacks can be simulated in a practical manner, manipulating key features of individual sample dataset elements to generate adversarial samples [52]. Furthermore, the study suggests dataset refinement strategies and improved learning models to bolster phishing detection against adversarial threats, ensuring greater reliability in real-world cybersecurity applications. The third study proposes an ensemble clustering approach to increase the resilience of NIDS against adversarial attacks [53]. By focusing on nonpayload connections at the TCP stack level, this research develops clustering-based data transformation techniques that improve intrusion classification. Unlike conventional wrapper methods, which rely on class label knowledge, this study introduces new filter models that operate independently of labelled data, making them more effective in identifying both direct and obfuscated intrusions. Collectively, these studies emphasize the crucial role of clustering in fortifying network security mechanisms.

4.2.3 Adversarial Learning with Clustering in Different Applications

This subsection is composed of six studies (N=6) that analyse adversarial learning with clustering in different applications. As personal decision-making becomes increasingly digital, the representation of users and assessment of trust are critical. The first study presents an adversarial fusion framework designed to make greater use of multiview information for user representation learning in social networks [54]. Most approaches do not successfully exploit multiview data or disentangle factors that determine user intention. By using adversarial learning, the proposed framework guarantees that user representations are comprehensive and robust to misleading data, which improves user personalization, recommendation, and security in social network contexts. The second study shifts the focus to trust management in industrial wireless sensor networks (IWSNs) by introducing a GAN-based mechanism to detect malicious nodes and enhance security performance [29]. The proposed framework strengthens the resilience of trust mechanisms in dynamic and adversarial industrial environments, ensuring reliable and real-time communication in sensor-based networks. In other domains, advanced clustering techniques in the fields of person re-identification (re-ID) and fine-art classification have been developed. These fields are situated in different contexts. The third study described in this section demonstrates that adversarial learning-based clustering methods can greatly enhance the discrimination of visual features in re-ID applications, and the fourth study focuses on the automatic labelling of fine arts paintings through clustering, thereby pushing the limits of machine pattern-directed recognition. The third study presents CANU, a conditional adversarial network to improve unsupervised person re-ID through clustering of visual features [55]. CANU offers the advantage of utilizing conditional camera adversarial training to improve the representation power of learned features during the clustering process. This shows that by improving identity consistency and removing artifacts from camera representations, the methodology greatly improves the accuracy of cross-camera person for re-ID applications, making it particularly useful for security and surveillance work. The fourth study focuses on fine art classification, proposing an adversarial clustering system (ACS) that enables the automatic labelling of paintings without human intervention [56]. Traditional fine-art classification relies heavily on manual annotation, which is often subjective and labor intensive. The proposed ACS model enhances unsupervised clustering quality by reducing the within-cluster sum of squares (WCSS) error and increasing classification accuracy in downstream supervised learning tasks. This innovation allows for the automated categorization of paintings based on machine-learned stylistic and compositional features. However, in other direction of application of industrial and fault diagnosis applications are needed. The fifth paper discusses the importance of adversarial learning for improved unsupervised clustering results. The authors introduced c-GCN-MAL, which is a clustering GCN with multiple adversarial learning for intelligent fault diagnosis in mechanical bearings [21]. In this context, traditional fault diagnosis systems struggle to process unlabelled data with considerable domain distance, which leads to a qualitative decrease in fault diagnosis power. The c-GCN-MAL combines adversarial learning to more precisely cluster data for improved fault detection model generalization ability and introduces a new loss function that aims to adapt the model to domain variations, in which the model can effectively transfer knowledge and improve a fault detection model with a previously unseen dataset. The last paper applies adversarial clustering to network representation learning by building the adversarial learning-based residual variational graph normalized autoencoder (ARVGNA) [27]. This study demonstrates the effectiveness of ARVGNA in critical graph-based tasks such as link prediction, node clustering, and graph visualization, particularly in challenging environments with isolated nodes or weak data representations. By incorporating adversarial learning, the model achieves more structured and informative embeddings, improving interpretability and predictive accuracy in industrial network systems.

4.2.4 Theoretical and Framework-Based Approaches

This subsection presents two studies (N=2) that develop theoretical propositions and explore adversarial learning frameworks with respect to clustering and network embedding. These ideas build on our understanding of how clustering can be manipulated and how adversarial learning works, creating new theoretical and real-world possibilities that change the way clustering models function and adhere to certain properties. The first study provides a plausible notion of an ε-semimetric where a mathematically objective way of assigning arbitrary distances between points in a dataset while minimizing violations of the triangle inequality [57]. This examination has important, useful conversations about ethics with respect to clustering algorithms in terms of how ε-semimetrics can be exploited to manipulate clustering, depending on subjective preferences. Certainly, ε-semimetrics can be used in a positive, nonadversarial way, for instance, increasing the usability or flexibility of clustering; however, they can also be imagined being manipulated in a fashion that distorts the data unethically, and conditional knowledge of distance-based clustering algorithms is reliable. The second study describes ArmGAN as confront adversarial learning framework for network embedding, with a perspective focused on how networks entail self-representation rather than later stages of drawing inferences based on embedding [58]. Unlike traditional GANs, which have been used for network representation learning, ArmGAN uses a three-player adversarial system composed of an autoencoder with mutual information regularization, a negative sample generator, and a discriminator, improving network embeddings and leading to learning latent representations that are more stable and informative when performing complex analyses of networks. The study evaluated ArmGAN and compared it to several network analysis benchmarks, and the results revealed that ArmGAN outperformed the state-of-the-art approaches in node classification, link prediction, and community detection. Both the first study and the second study were influential in addressing the philosophy and the technical aspects around adversarial clustering and representations learning. Thus, both studies add to the theoretical development of machine learning applications in clustering-based applications that are more resilient, interpretable, and ethically speaking to their approaches.

4.3 Based on RSs and Adversarial Learning

This section explores the complex relationships between recommendation systems and adversarial learning, highlighting their intersection as a pivotal area of study. Through an examination of seventeen scientific studies (N=17), we analyse the impact of adversarial strategies on the stability and effectiveness of recommendation systems. The following subsections present a systematic classification of these studies, offering a structured understanding of the key approaches and advancements that arise from this intersection.

4.3.1 Poisoning and Injection Attacks in RSs

This subsection consists of seven studies (N=7) that refer to adversarial strategies where malicious entities manipulate input data to distort recommendations. These attacks can degrade system accuracy, promote deceptive content, or undermine user trust. Poisoning attacks corrupt training data, whereas injection attacks introduce fake profiles or interactions to bias outcomes. The first study examined the GraphRfi recommender system and suggested the GraphRfi can become vulnerable to node injection attacks, where malicious actors create fake profiles to manipulate the recommendation process [23]. The study utilized an advanced attack method called MetaC to reveal how to exploit the recommendation system; however, rather than normalizing being vulnerable to link injection attacks as an effect of openness, the study also suggested a dynamic-scale based adaptive fraudster detector that revisits the adjustments made to the newly added user, which led to the development of PDR: a protection framework for learning systems that combines anomaly detection with the learning framework, the learned framework should create a recommendation process that allows it to be completely robust to adversarial acts. The second study extends the arguments to conceptualize a comprehensive framework for detecting shilling attacks while also engaging the question of how fraudsters operate in recommendation systems [59]. By examining the evolution of user and item embeddings before and after attacks, the stealth and effectiveness of adversarial strategies can be investigated. This study provides critical insights into the perpetual arms race between recommender system security and adversarial ingenuity. In contrast, the third study presents the triple cooperative defense (TCD) approach to increase the robustness of a recommender system against poisoning attacks and presents poisoning methods through the introduction of co-training attack (CoAttack) and game-based co-training attack (GCoAttack) methods to maximize attack efficiency in a cooperative training setting [60]. The fourth study develops poison-tolerant collaborative filtering (PTCF), a method intended to allow a CF recommender system to continue to function despite poisoning attacks [61]. PTCF represents a departure from preparing for security events, allowing CF to occur on a poisoned dataset while being resilient to the impact of data poisoning on system availability and functionality. Additionally, the fifth study takes a dual approach by both investigating the vulnerabilities of the RSs and proposing countermeasures. It introduces InfMix, a poisoning attack strategy that employs an influence-based threat estimator and a user generator to construct malicious profiles, effectively testing the system's susceptibility to manipulation [62]. In response, the study develops adversarial poisoning training (APT), a defense mechanism that proactively injects synthetic users designed to minimize empirical risk and reinforce system robustness. . Moreover, the sixth study reported here is DSSD-ImMPL, a new detection method that targets the identification of many attacks in recommendation system datasets [63]. There are particular interests in classical and mixed attacks that make them difficult for some detection frameworks to manage. The seventh study is the KC-GCN, which describes a two-stage semisupervised detection model developed to engage in group shilling attacks [64]. Group shilling attacks exploit large groups of users collectively manipulating the user profile and notably conflict with the detection frameworks because of the compositional complexity related to varying adversarial and mixed attack groups. The KC-GCN model addresses a significant gap in attack detection methodologies by focusing on this vulnerability, offering a more robust solution.

4.3.2 Adversarial Learning for RSs Robustness

This subsection discusses three research studies (N=3) that employ adversarial learning techniques to increase the robustness of RSs. These studies explore different recommendations for how to respond to adversarial threats, enabling more reliable and secure recommendations. The first study proposes stagewise hints training and randomized noise layers that together improve the resilience of recommendation models while maintaining predictive accuracy [65]. This technique provides an opportunity for learning that is robust and resilient to adversarial manipulations. The second study defines DAAN, a novel cross-domain recommendation framework that utilizes matrix factorization CF with deep adversarial domain adaptation [66]. DAAN uses an attention network to weigh strategies for effectively balancing domain-shared and domain-specific strategies and to address issues of data sparsity. The resulting recommendation model was proven to increase both the robustness and accuracy of the recommendations. These studies enter the literature to reinforce RSs from adversarial threats. The third study explores various malicious user detection algorithms and introduces a novel framework designed to enhance detection performance [22]. By refining representation learning techniques, the proposed approach strengthens the accuracy and efficiency of detection systems, ensuring more reliable identification of deceptive behaviors.

4.3.3 Adversarial Attacks on Graph-Based Recommendation

This subsection examines five studies (N=5) related to adversarial attacks on graph-based recommendation systems. The first study presents (CopyAttack+), a reinforcement learning-based framework that uses combinations of cross-domain user profiles to mount black-box adversarial attacks. (CopyAttack+) establishes a local surrogate system and trains it to improve attack success and improve the adaptability of the attack to the weaknesses of the black-box recommender system [25]. The second study examines knowledge graph-based recommendation systems under poisoning attacks in which fake links are inserted to influence recommendations. The focus of an attack such as this is to increase the visibility of tagged products and influence recommendations without being easily detected [67]. The third study investigated adversarially learned injection attacks in graph-based recommendation systems [3]. Researchers have proposed a better detection method to thwart advanced adversarial attacks against recommendation systems that would otherwise evade traditional defenses. The attackers of these systems use knowledge graphs, and there are ways that can leverage knowledge graphs to fortify structures originally used for the detection of advances and against adversarial injections made against these recommendations. With this methodological improvement, we can better understand harmful behavior, with the aim of improving detection and therefore improving graph-based RSs against an adversarial attack. The fourth study is a graph-contextualized trip recommendation, GC-TripRec, which improves trip recommendations produced through adversarial learning by identifying and capturing more complex relationships regarding point-of-interest (POI) [68]. This study uses graph representation learning to create trip recommendations in conjunction with POI global and trip representations naturally. The model provides better contextual understanding regarding user preferences, better adaptation to the POI to user preferences across POI during the trip, and ideally improves the user experience by providing better and more relevant trip recommendations. The fifth study investigates how hyperactive users impact political dialogue across online social networks (OSNs) on social media, and relationships draw out similar benefits, making viewpoint statements or viewpoints dominate along with ways to affect some factors incidentally through bias [69]. By examining the role of adversarial attacks on recommendation algorithms to expose or suppress particular content or information, they amplify biased narratives to reconstruct the perceived political realm. Their results suggest that adversarial attacks manipulate recommendation systems to deploy and steer modes of discussion, ideology, and negligence.

4.3.4 Deep Learning-Based Attack Detection for RSs

This section presents two studies (N=2) that examine advanced deep learning-based techniques for discovering adversarial attacks in recommendation systems. The goal of these studies is to improve the efficient use of models (e.g., performance, accuracy, and precision). The first study develops CNN-BAG, a new hybrid method that combines convolutional neural networks (CNNs) with bagging (BAG) for the discovery of recommendation attacks. CNNs are ideal because they are deep neural networks that automate the process of feature extraction, reduce human adjustment, and improve detection performance when confronted with attacks [2]. The second study examines the complexity challenges of generalized adversarial network recommendation models, which is a barrier to their recognized applied use [70]. This study focuses on parameter efficiency, knowledge transfer, and model compression, with a special emphasis on the difficulty of optimizing student models when trained on adversarial training data. This is important because, in adversarial training, noisy knowledge can affect the performance of attack detection systems. These studies contribute to deep learning-based adversarial detection, which specializes in efficiency and readability when executing applied use in particular.

4.4 Integration based on RSs, Clustering, and Adversarial Learning

This section of this paper (N=4) analyses the integration of RSs, clustering, and adversarial learning and identifies the value towards the development of more intelligent and efficient recommendations. RSs are the basis of personalized content delivery that dynamically adapts to a user's preferences in order to improve engagement. However, as data become increasingly complex and voluminous, there is a greater demand for and need for structured approaches. Clustering affords models an efficient means of organizing a set of users or items by any inherent similarities in data instances to improve the accuracy and computational efficiency of the recommendations. This is critical for system scalability. Adversarial learning through the use of reinforcement learning methods enables greater diversity and stability of RSs by being able to account for shortcomings of the learner (such as indications of poisoning and faulty recommendations). Clustering, RSs, and adversarial learning are complementary and incorporate three critical aspects of AI that support a more intelligent, secure and scalable recommendation model and are valuable and highly sought after in the area of research for next-generation AI-driven personalization systems. Three subcategories of the analyses of the targeted studies are addressed in the following analysis.

4.4.1 Improving RSs

This subsection provides a total of two papers focusing on (N=2) RSs, which are essential for filtering large pools of data and the importance of providing users with effective suggestions, with the importance of improving their accuracy, efficiency, and resiliency. To improve RSs, new techniques must focus on improving user interaction modelling and learning strategies [71]. One paper proposed a new bandit problem called Online Learning and Detecting Corrupted Users (OLDCU), which is used to infer latent user relationships through online behaviors as they dynamically evolve. Using a conversational learning approach, this study proposes analysing user interactions with the RCLUB-WCU bandit algorithm in order to develop effective relational inferences. The paper also proposed a detection algorithm, the OCCUD framework, in order to automatically learn and refine a set of possible user patterns over time for continual improvement of recommendations. Furthermore, another paper attempted to develop an unsupervised divide-and-conquer method to classify profiles in RSs [72]. This method proposes a distinction between standard attacks and obfuscation behavior attacks by developing a separate model for each type of classification. Overall, this method removes the traditional aspect of needing any annotated references, allowing it to autonomously, and on a scalable level, identify any sort of behavioral pattern. In summary, both studies classify and improve the recommender system through computational efficiency to help maintain the integrity of a recommender system and therefore provide a system with the ability to produce high-quality personalized recommendations.

4.4.2 Dealing with Fake Data and Attacks to Robust Performance

This subsection is based on one research paper (N=1) in which ensuring the resilience of an RSs against adversarial manipulation necessitates a comprehensive understanding of how synthetic user profiles influence system performance. One study critically examines the ramifications of single-user adversarial control, wherein an attacker operates with an extremely limited number of fake users (potentially as few as one) to compromise the integrity of recommendation outputs [31]. To formalize this phenomenon, the study introduces a clustering-based framework for generating synthetic user profiles, which can be strategically deployed within poisoning attacks targeting deep learning-driven recommender architectures. By demonstrating the efficacy of these attack mechanisms, this study underscores the imperative for more robust defensive strategies, suggesting the development of countermeasures capable of mitigating adversarial perturbations at both the structural and algorithmic levels. Through this lens, the study contributes to a broader discourse on fortifying recommendation models against data-driven manipulations, ensuring sustained performance reliability.

4.4.3 Recommendation in Distributed Environments

This subsection is based on one research paper (N=1) in which the evolution of RSs in large-scale, heterogeneous infrastructures necessitates the development of distributed learning paradigms that enhance efficiency and adaptability. One study introduced the distributed variational autoencoder sequential recommendation method (DistVAE), a distributed variational autoencoder designed to optimize sequential recommendation processes within decentralized computational frameworks [32]. By leveraging the availability of diverse and distributed infrastructures, DistVAE aims to refine recommendation accuracy while maintaining scalability across complex environments. A key methodological innovation in this study is the mitigation of gradient randomness during distributed model aggregation. To achieve this, the research employs the Gaussian mixture model (GMM) clustering algorithm, which systematically stabilizes gradient variations across multiple computational nodes. This methodological refinement ensures that recommendation models trained in distributed settings can achieve greater convergence stability and predictive reliability, ultimately fostering robust, scalable, and high-performance recommendation architectures within decentralized environments.

5. DISCUSSION

This section aims to elucidate and discuss three fundamental concepts derived from the gathered articles: (1) the motivations, advantages, and significance of the issues that prompted researchers to emphasize and seek solutions for problems; (2) the challenges encountered by current and former researchers regarding the cases and obstacles reported; and (3) the recommendations and prospective work suggested by the authors concerning future applications at the intersection of RSs, clustering, and adversarial learning.

5.1 Motivation

This section provides a comprehensive analysis of the underlying motivations derived from the extracted studies. Figure 7 illustrates that these motivations have been systematically categorized into six primary groups, each reflecting a distinct aspect of the intersection between adversarial learning, RSs, and clustering methodologies. A detailed discussion of these categories is presented, highlighting the interdependencies that shape advancements in these fields.

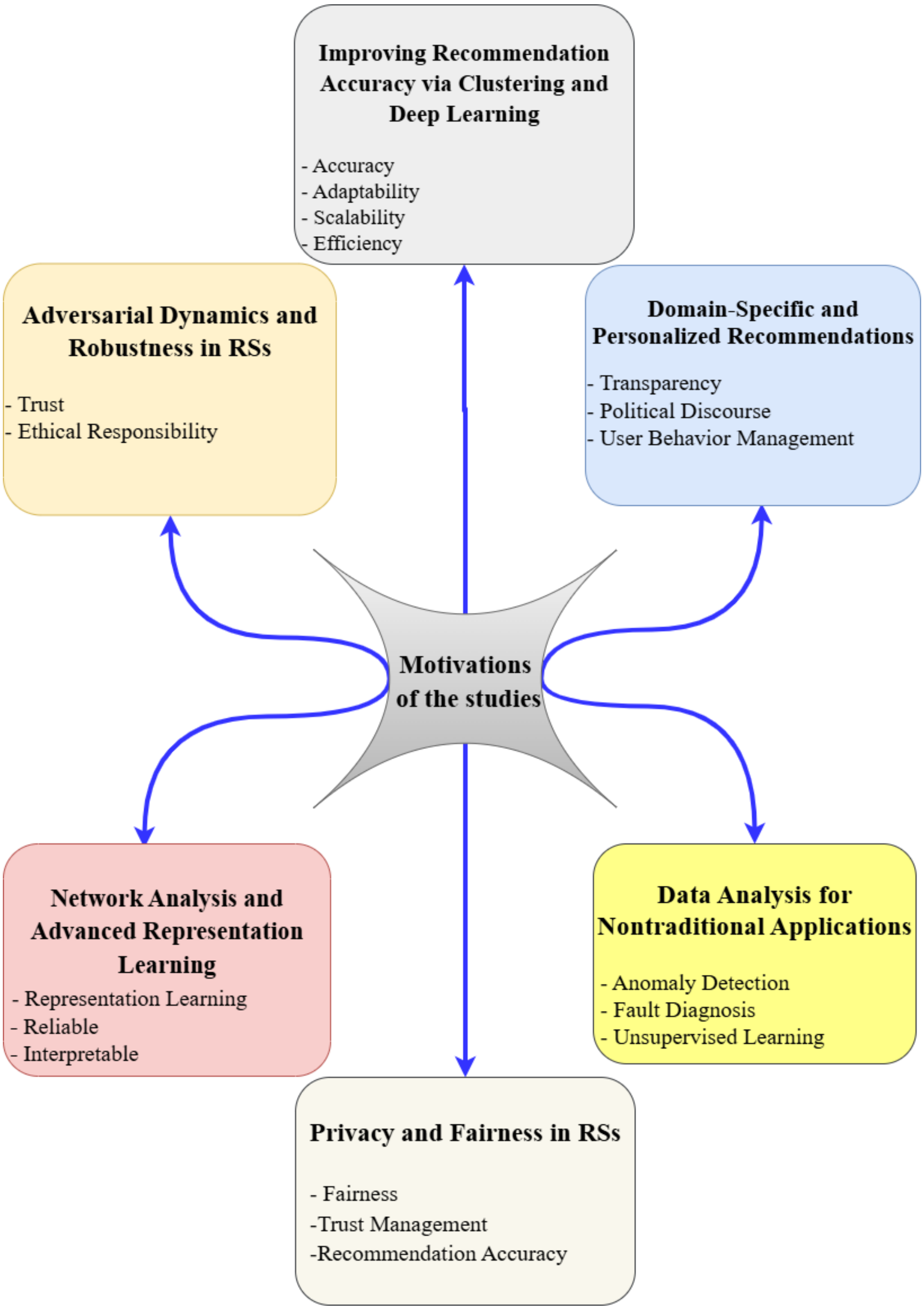
Improving Recommendation Accuracy via Clustering and Deep Learning
- Accuracy
- Adaptability
- Scalability
- Efficiency
Adversarial Dynamics and Robustness in RSs
- Trust
- Ethical Responsibility
Domain-Specific and Personalized Recommendations
- Transparency
- Political Discourse
- User Behavior Management
Motivations of the studies
Network Analysis and Advanced Representation Learning
- Representation Learning
- Reliable
- Interpretable
Data Analysis for Nontraditional Applications
- Anomaly Detection
- Fault Diagnosis
- Unsupervised Learning
Privacy and Fairness in RSs
- Fairness
-Trust Management
-Recommendation Accuracy

Fig. 7 Motivations of the studies

5.1.1 Adversarial Dynamics and Robustness in RSs

The drive to protect RSs from malicious attacks stems from a blend of practical, ethical, and strategic imperatives [46]. At its heart is a fundamental challenge: balancing the openness needed for personalized recommendations with the security required to maintain trust [67]. Studies such as MetaC and InfoAtk reveal how attackers exploit weaknesses in systems designed to detect fraud (e.g., GraphRfi), turning their own logic against them [59]. Rather than viewing these flaws as dead ends, researchers use them as starting points to innovate defenses such as PDR, which adapts to uncertainty, proving that vulnerabilities can fuel progress when approached with creativity [23], [27] Central to this effort is redefining how systems identify threats. Tools such as Infmix and CNN-BAG address the blurred line between legitimate user behavior and hidden attacks [2], [62]. Ethically, the focus shifts from chasing perfection to managing imperfection. Methods such as PTCF and TCD accept that no system can fully eliminate poisoned data [60], [61]. Instead, they build resilience by working with compromised datasets, acknowledging that real-world data are often flawed [3]. This pragmatic mindset extends to societal trust studies such as ClusterPoison and Unsupervised Contaminated User Profiling, which highlight how attacks on recommendations are not just technical breaches that erode user confidence. ensuring that users retain control over algorithmic decisions [31], [72] The back-and-forth between attacks and defenses drives progress. Each new threat, such as poisoning clustering algorithms, sparks better solutions, creating a cycle where systems evolve by confronting challenges head-on.

5.1.2 Improving Recommendation Accuracy via Clustering and Deep Learning

The motivation for these studies was based upon the increasing need to improve the content accuracy, efficiency, and adaptability of RSs in diverse contexts [64], [68]. An increasing digital landscape is characterized by an increase in applications that stretch across e-commerce and online education, smart cities, and the IoT [38], [70]. It is becoming even more critical to provide accurate and customized recommendations. These studies highlight the importance of using sophisticated methodologies to refine recommendation approaches, providing more relevant interactions with users and better decision making [1], [4], [39]. The integration of clustering methods and deep learning methods supports the identification of complicated schemas that determine user behavior, preferences and related contextual factors to increase the accuracy of recommendations [42]. Additionally, acknowledging the integration of generative models and knowledge distillation supports a better representation of the data and minimizes complexity, which could yield data that scales and are deployable [41]. Collectively, these studies look to enhance RSs towards personalization, fairness and adaptability, which can enhance user experiences across digital ecosystems.

5.1.3 Domain-Specific and Personalized Recommendations

The impetus for these investigations arises from the volatile and evolving landscape of domain-related and personalized recommendation systems, which cater to unmatched demands relating to distinct areas across a disparate range of domains such as political discourse, news communication, movie recommendations and user behavior management on digital platforms. One of these studies examines the hyperactive use of political communication in OSNs, particularly how the disproportionate influence of hyperactive users influences public opinion and affects the outcome of their recommendations [69]. The study advocates algorithmic transparency in order to promote fairness and democratically empower users. The second study sought to develop a recommendation system based on news, utilizing the semantic richness of the news content, in order to enhance its ability to provide a more comprehensive feature and varied recommendation model that would improve user engagement and satisfaction [37]. The third study develops a recommendation system for movie recommendations that sends content-based recommendations for discovering movies that ensure that they match their personal preferences [43]. The fourth study addresses how to improve recommendations for “new” users using both content and collaborative recommendations for initial recommendations to persuade the user to trust and engage with their recommendations even when the user has had no prior sign of activity [40]. The fifth study outlines optimization strategies using collaborative signals such as knowledge graphs and CF to enhance item representation and user representation [44]. Finally, the sixth study takes a new angle to reward user behaviors such as misleading ratings by declaring and leveraging social connections and user relations to enhance the intelligence and speed of users [71]. All six studies contribute to the body of literature on personalized recommendation systems that seek to increase their accuracy, with fairness and user satisfaction across different domains.

5.1.4 Network Analysis and Advanced Representation Learning

Network analysis and representation learning have progressed to the point that researchers are further focusing on methods with more complex models that can encode and learn complicated relationships within data [66]. The same goal of improving user representation learning, knowledge transfer, and multiview information should be adopted, thus improving the transferability and intelligibility of machine learning systems [48]. A primary area is continuing to improve cross-domain recommendations that utilize both types of shared knowledge from the source domain, but the model also learns, doing this with risks presented explicitly in the source domain and not requiring data from directly from topically similar domains [47]. Studies of clustering algorithms also suggest that they can be extended to incorporate both topological and feature information, and findings suggest that dual linages have the potential to be more meaningful than embedded clusters [54]. The contributions of generative and adversarial learning include infinitely improving the expressiveness of latent representation spaces, potentially when the probability convergence of embeddings is stopped, thus enabling social networks, RSs, and graph-structure-based learning [58]. These "commensurate" studies working to increase the robustness of network embeddings and deep clustering algorithms with improved architecture aim to contribute to more efficient, portable, reliable, and interpretable network analysis [5], [27]. The improvement of adversarial learning and deep generative methods has led to enormous improvements in representation learning methods, and this progress can also be seen in better decision making with recommendation systems and better information retrieval models that leverage network structures, each hopefully from sufficiently wide distributions of applications.

5.1.5 Privacy and Fairness in RSs

The driving context for each of these studies is the improvement of privacy, fairness, and trust in the RSs across multiple applications. One study focused specifically on fairness and objectivity in clustering algorithms when adversarial manipulation was present and not well rooted [57]. It investigates how distance functions can be designed to produce unbiased results, which ultimately increases the trust of users in any cluster assignment. A second study is inspired by the need to promote healthy market practices in smart cities and improve analyses of large volumes of multidimensional market data [49]. This study proposes advanced techniques, eClusterGAN and GAN intrusion detection systems to improve clustering accuracy and data storage safety. A third study is guided by the goal of improving trust management in IWSNs and employs GANs to improve (i) resilience and (ii) adaptability [29]. A fourth study seeks to improve e-commerce RSs via the generation of recommendations that account for trust, user overlap, and time to improve the accuracy of recommendations [19]. A final study aimed to improve recommendation system performance via DistVAE, which captures long-term dependencies in user interactions to improve accuracy [32]. This aspect was a focus of the final study. Each of these studies has an aligned goal in improving privacy, fairness, and trust and hence building better and more accurate RSs.

5.1.6 Data Analysis for Non-Traditional Applications

The increasing complexity of real-world data and changing security threats has pushed research toward advanced analytical techniques and applications that extend beyond traditional use [55]. A common impetus across several areas of research is to improve the performance of machine learning models at recognizing anomalies, recognizing adversarial threats within data, and finding subtle patterns in data from complex environments [63], [65]. These areas of research range from improving the robustness of recommendation systems from adversarial attack or manipulation to improving network intrusion detection and phishing prevention mechanisms; the goal is to improve the analytics in data-driven defenses [22], [45]. More recently, with advances in unsupervised learning, it has become possible to improve the recognition and classification of complex entities in limited instances with little expert labelled data, such as individuals in surveillance video footage or paintings in fine art [56], [58]. Research has also focused on learning meaningful representations in adversarial situations with a point of emphasis on clustering techniques and the ability to adaptively cluster, where finding the optimal number of clusters does not rely on the subjective choice and can even be determined dynamically with improved accuracy classification [50], [53]. Thus, in the industrial or organizational realm, there is a need for robust fault diagnosis methods that are resilient to domain shifts to provide accurate detection of failure [21]. Overall, these aspects of research can be seen as contributions to the more general conception of data analytics across unconventional fields and in developing methodologies that can search for the right mixture of interpretability, adaptability and security in their machine learning objectives [52].

5.2 Challenge

This study examines the challenges encountered by researchers at the intersection of RSs, clustering, and adversarial learning. These challenges have been systematically categorized into seven distinct clusters, with each cluster comprising studies that address similar issues. The classification framework, as illustrated in Figure 8, provides a structured overview of the shared research obstacles within these domains.

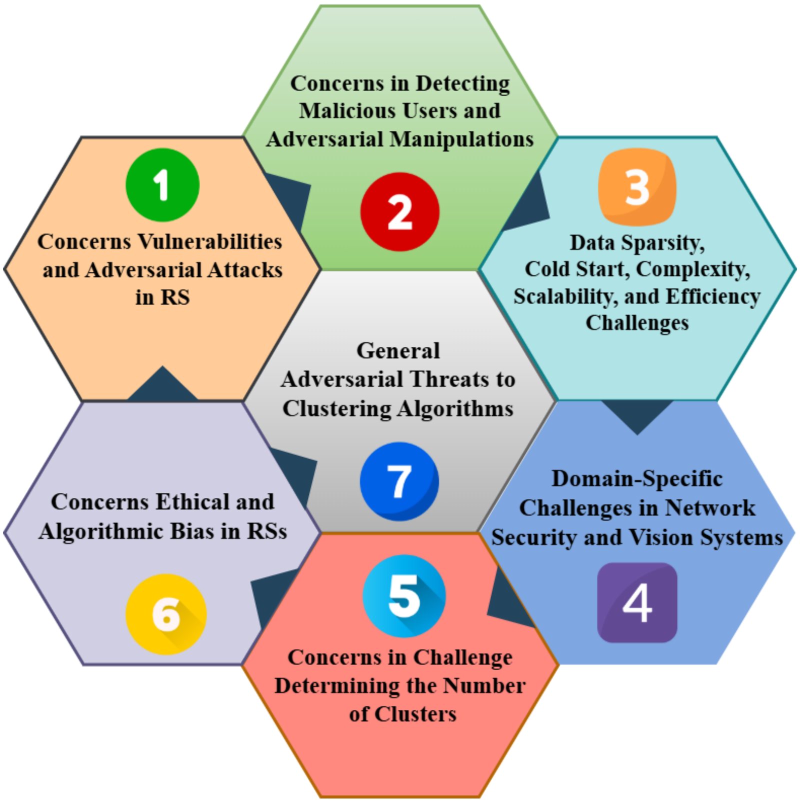
Concerns in Detecting Malicious Users and Adversarial Manipulations
2
1
Concerns Vulnerabilities and Adversarial Attacks in RS
3
Data Sparsity, Cold Start, Complexity, Scalability, and Efficiency Challenges
General Adversarial Threats to Clustering Algorithms
7
Concerns Ethical and Algorithmic Bias in RSs
6
Domain-Specific Challenges in Network Security and Vision Systems
4
5
Concerns in Challenge Determining the Number of Clusters

Fig. 8 Challenges of studies

5.2.1 Concerns Vulnerabilities and Adversarial Attacks in RSs

The increasing reliance on RSs across various domains has raised significant concerns regarding their vulnerability to adversarial manipulation. These challenges manifest in multiple forms, including susceptibility to fake user profiles, poisoning attacks, and complex optimization problems inherent in both attack strategies and defense mechanisms[5], [61], [62]. One fundamental difficulty lies in the supervised nature of fraud detection, which depends on the availability of clean labels and struggles against sophisticated adversarially learned attacks that mimic genuine user behavior [3]. Additionally, existing defense strategies face limitations in terms of generalizability and robustness, particularly when confronted with bi-level optimization problems that intertwine attack adjustments with model parameter updates [60].

CF systems are particularly sensitive to adversarial perturbations, with their performance degrading significantly under malicious interference [65]. While knowledge graph-based recommendation systems are inherently susceptible to adversarial threats, the overwhelming number of possible ways to attack the system further complicates this vulnerability; it is impractical to analyse and defend against all potential attacks on a knowledge-based recommender system. In addition to threat vulnerabilities, reinforcement learning-based RSs are challenged by the need to manage large discrete action spaces, develop continuous and accurate profiles across task domains, and utilize useful transferable signals to improve adversarial learning strategies [67]. In today's systems, learning the accuracy of the modelled user representation is also a hinderance, as many existing models do not disentangle the many complex layers of interactions between the user's higher-level and more specific intentions, resulting in weak interpretability and nonrobust user representations [54]. Even further still, it is difficult to accurately assessing the potential harm caused by adversarial acts, particularly in bi-level poisoning, which is a crucial factor in evaluating the threat, is difficult [25]. These vulnerabilities demonstrate the necessity for flexible, robust, and interpretable RSs capable of withstanding adversarial actions without compromising system performance and fairness. To mitigate these issues moving forward, a shift in orientation should occur, defensively with a focus on multiview learning, adaptive user profiling, and adversary-aware optimization strategies.

5.2.2 Concerns in Detecting Malicious Users and Adversarial Manipulations

The ever-increasing sophistication of adversarial manipulations of RSs and networks is hindering detection methods. A malicious actor can take advantage of the design flaws in these systems to generate fictitious user profiles, generate fictitious ratings, and bypass detection methods by using attacks based on deep learning, such as GSA-GAN and adversarially learned injection attacks [3], [63]. These types of attacks are designed to closely resemble genuine users to confuse and overwhelm standard input detection models that are trying to determine whether the user or the entity is legitimate. The overall performance of the detection models is significantly weakened when there are insufficient labelled training data, and the decline in overall accuracy is further complicated by the discovery of novel attack schemes. Furthermore, group shilling attacks pose additional challenges by implementing the GOAT adversarial techniques or mixed attack groups. It is especially difficult to detect smaller group shilling attacks, as the models we have been developing say they can be easily fooled when the malicious number of participants are reduced [64]. Additionally, hyperparameter tuning interferes with stability in detection and limits the accuracy of the results, making them very sensitive to any changes in parameter selection. Like identifying influential users within a network graph, detecting attacks is a very difficult analytical task that may be assumptive to simple statistical descriptions without fully considering how online interactions progress. We therefore need detection approaches that are adaptive to new and emerging attacks. One primary challenge in detecting malicious users arises from the use of traditional collaborative filtering recommender systems (CFRSs), which rely on user ratings to generate recommendations [3]. Sophisticated attackers can generate fictitious ratings that may conceal or bias recommenders. The situation is worse, as attack profiles are often underrepresented (although they outnumbered user profiles), leading to detection methods failing in an absolute sense but performing poorly with supervised methods. Like others, deep learning or learned-based recommendation attacks, such as GANs, are very strong at concealment and are not conducive to detection methods that depend on features derived from user profiles [2]. Importantly, we have an automatically adaptive detection approach that learns not only the attack space but also the relationship between the adversarial attack profile and exposure. A further challenge in the area of recommendation security is detecting corrupted user behaviors. When a web-based recommender system has deceptive click ratings or artificially manipulated interactions, traditional bandit learning algorithms can be disrupted significantly because the nature of the algorithms developing learning knowing only about the individual user and failing to consider their implicit social relationships with other users [71]. There is a pressing need for novel approaches that can dynamically infer user relationships and detect corrupted behaviors in real time. Moreover, existing profile identification methods are often tailored to specific attack scenarios, limiting their generalizability. Many require labelled data, leading to high annotation costs and risks of overfitting. An unsupervised approach that can be generalized across different types of attacks, particularly standard and obfuscated behavior attacks, would greatly increase detection precision [72]. Another key challenge in adversarial detection involves generating realistic malicious user profiles for training robust models. Ensuring that synthetic users accurately reflect the distribution and diversity of real malicious users remains difficult [22]. Traditional data augmentation techniques fail to balance these aspects, leading to inconsistencies that weaken the detection effectiveness. Similarly, GAN-based network embedding methods struggle to distinguish between meaningful node representations and Gaussian noise, compromising overall performance [58]. The existing adversarial learning strategies apply primarily to representation results rather than the mechanisms themselves, limiting their ability to capture the full potential of GANs. In addition to recommendation systems, adversarial manipulations extend to cybersecurity threats, such as phishing and network intrusion detection. Attackers must carefully modify phishing website features to evade classifiers while preserving functional and visual coherence, increasing detection risk. Minimizing both the number of manipulated features and the cost of modifications remains a challenge, as excessive changes increase the likelihood of exposure [52]. Similarly, adversarial attacks on NIDS pose difficulties in classifying legitimate traffic, direct attacks, and obfuscated intrusions [53]. Mutated traffic patterns further complicate detection, as they often resemble normal network activity. The complexity of generating effective base clustering for intrusion detection models presents additional constraints, particularly in resource-limited environments. In light of these challenges, the development of more robust adversarial detection mechanisms is imperative. Given the rapid evolution of attack techniques, detection frameworks must not only identify malicious behaviors with high precision but also anticipate and adapt to emerging threats in an ever-changing digital landscape.

5.2.3 Data Sparsity, Cold Start, Complexity, Scalability, and Efficiency Challenges

The effectiveness of RSs is increasingly limited by issues such as data sparsity, cold start problems, scalability constraints, computational complexity, and efficiency limitations. These challenges arise from the rapid expansion of data in digital environments, the limitations of traditional recommendation models, and the growing demand for personalized recommendations with minimal computational overhead [1]. Addressing these concerns requires advancements in hybrid modelling, knowledge distillation, clustering techniques, and distributed learning methodologies to ensure more accurate, scalable, and efficient recommendation frameworks. Data sparsity remains a fundamental issue, as users typically interact with only a small subset of available items, leading to insufficient data for reliable recommendations [19], [42], [43]. This limitation significantly affects CF methods, which rely on shared user–item interactions [39]. The problem is further compounded by the cold start issue, where newly introduced users and items lack historical data, making it difficult to generate personalized recommendations. To mitigate these challenges, hybrid recommendation models that integrate content-based and CF techniques have been proposed, allowing for a more robust feature representation and improved adaptability to new data [4]. Another critical challenge is the high computational complexity of deep clustering and deep learning-based recommendation models[70]. The training of complex models, such as eClusterGAN and deep variational autoencoders, often requires extensive resources, making their application in real-world scenarios difficult [49]. The dependency on predefined cluster numbers and the instability of generative adversarial training further hinder the practical deployment of these models. Additionally, high execution times in recommendation processing can negatively impact user satisfaction, necessitating optimizations in clustering efficiency, feature extraction, and representation learning [40]. Scalability poses yet another significant barrier, particularly as the volume of user-generated data continues to grow exponentially [41]. Traditional recommendation techniques struggle with the increasing number of potential neighbors in CF approaches, making real-time processing inefficient. The complexity of computing the similarities between users and items further exacerbates this issue, especially in IoT-driven environments where millions of data points must be processed simultaneously [39]. To address this, distributed learning approaches such as the DistVAE model leverage heterogeneous infrastructures to increase computational efficiency while improving recommendation accuracy through client-based clustering [32]. In light of these challenges, future advancements in RSs must prioritize the development of efficient, scalable, and interpretable models. By integrating hybrid techniques, optimizing clustering strategies, and leveraging distributed learning, researchers can overcome the limitations of existing methods. The continuous refinement of these approaches will be crucial in ensuring that recommendation systems remain effective, adaptable, and capable of delivering high-quality, personalized recommendations in increasingly complex digital ecosystems.

5.2.4 Domain-Specific Challenges in Network Security and Vision Systems

Clustering techniques face certain limitations in domain-specific set-ups in network security and vision systems. More specifically, IWSNs, where there are limited labelled examples for novel attacks and trust uncertainty that occurs because of node failures, inhibit our potential to harness previous experiences to become more informed on anomaly detection. The limitations of the sensors themselves make creating trustworthy models or even robust trust models challenging [29]. Similarly, unsupervised person re-ID faces challenges with angle view variations, where images are clustered together of persons who are verified to look similar because of camera view rather than identified persons. This issue is exacerbated by the negative transfer and usage of pseudo-labels. It was proposed that conditional adversarial networks have been proposed to be more efficient at clustering [55]. Both areas highlight the need to develop better adaptive techniques to address domain-specific constraints regarding security and vision systems.

5.2.5 Concerns in Challenge Determining the Number of Clusters

The problem of choosing the number of clusters creates a serious problem in clustering methods, particularly with respect to fundamental difficulties with deep spectral clustering. A substantial problem arises from the use of ε-semimetrics, which can skew distance calculations in a subtle way with the appearance of objectivity, and as a result, unethically clustered solutions may be obtained [57]. The manipulations made to distance could be considerably more risky when compared to normal data poisoning attacks; most non-experts are not aware that they have been manipulated with respect to the distance metric, creating issues of fairness and reliability in clustering as a basis for decision-making.

Conventional spectral clustering has issues with leveraging the graph structure to its full potential, but with deep spectral clustering, the problems escalate, and the clustering performance is compromised [50]. The challenge is very complicated, as we have noted, and while we can optimize deep spectral embeddings, there is not a one-size-fits-all approach; one would need to determine specific methods for specific datasets, making the issue of clustering challenging across a diversity of applications. If not trivial, hyperparameter tuning is an ongoing concern. We know that if we can tune the hyperparameters of the model, we will increase how accurate the clustering is, but it is not easy, especially if all we have real datasets for which we do not know the ideal configurations. Addressing the set of questions and challenges raised in this segment requires developing robust, scalable, and interpretable clustering techniques that can accommodate the historical and structural characteristics of the datasets while maintaining computational efficiency and integrity and are performed without explicit bias.

5.2.6 Concerns Ethical and Algorithmic Bias in RSs

The increased use of RSs to influence social (digital) discourse and its potential ramifications with respect to commitment to ethics and algorithm dynamics has garnered robust coverage [21]. These systems drive user preferences and decision-making; however, their inability to provide transparency and political contextual intelligence may pose serious ramifications. In this context, a salient issue remains regarding recommendation algorithms that control political communication in online social networks. The lack of clarity in data-dominant RSs restricts the understanding of the role algorithms play in shaping political discourse, especially when formalized knowledge can be deduced from algorithm mediation, thus potentially altering the discourse on political theory, political discourse, collective action, and social movements [69]. This challenge is further compounded from hyperactive users that tend to bias the recommendation from data that are driven from recommendations that they may over-influence, which can distort public debate from large and disparate user engagement. These issues call into question the recommender system's role as a political mediator, especially from the perspective of expanding the equality and fairness of algorithm-produced outputs. In addition to the political discourse effects of RSs, structural weaknesses, such as deep multiview clustering architectures, exist in RSs. For example, these systems not only are limited in their robustness but also often lack the physical patterns and experiential representation of many views/channels of their actual semantics [48]. These common weaknesses are consequential and exasperated by adversarial attack strategies, as machine learning-based RSs have anatomical weaknesses in their algorithms that users can exploit to induce false outputs. For hypotheses and experimental news RSs, structural weaknesses extend to problems of timely recommendations and recommendations that can affect user personalization, which can be severe in systems where the need to recommend and action is constant and frequently changing. Timely recommendations tend to be a relevant problem due to the news cycle, which makes news rapidly obsolete with infrequent usage, inducing cold start problems or seriously limiting an evolving understanding of user interest and preferences [37]. The lack of obvious user feedback makes personalization even harder and suggests that some types of recommendation systems need to be clear and dynamic. Clearly, algorithmic bias is a major obstacle for researchers, especially for subjective discretion pathways such as in some aspects of artistic classification [56]. Similarly, ethical issues exist for consideration in health, and recommendation systems are often murky; defensibility and similarity measurements plus issues with sparseness can suggest too much and therefore detract from purpose and greater action. Although demographic elements such as age and health for recommendation engines may mitigate bias, they all too readily collapse any distinctions in terms of rights or privacy. We also note that the convergence of RSs and cybersecurity presents weaknesses. NIDS can be presented with class imbalance and bias risk through successful attempts at adversity with classes [51]. Furthermore, traditional over-sampling methods can easily create overfitting errors, whereas under-sampling methods can create meaning but additional computational costs. For example, in phishing detection, suggestions we concede in creating adversarial samples that are classifiable without loss of look or functionality from concerned samples. This approach also captures the risks that exist for adversarial manipulators while at the same time making broader appeals against algorithmic safety: confidence in future work on solutions in recommendatory structures. These issues are also similar for models relying on variational autoencoder-based models, which have been explored to improve performance and struggle with biased variational inference, underfitting, and inference gaps due to data sparsity [5]. Adversaries may have caused concerns that unique and minimally injected profiles can reasonably effectively compose an adversary to the base norms of not only the RSs but also suspected claims that might coexist with false users to perpetrate a faux attack on a recommender system. In closing, to whatever extent the entitlement of any guidance to manage ethical and algorithmic bias, as initially presented in RSs, is possible, some ranging methods will work. Future use of any works in these areas must demonstrate how to be distinct and open with systems, and there are clear definitions for adversarial defenses and any automated method for ethics assessment. These findings present vital work ahead both in research and practice in terms of how we work ahead.

5.2.7General Adversarial Threats to Clustering Algorithms

Clustering algorithms are increasingly confronting new challenges that impede resilience against adversarial threats, which may hinder the robustness of clustering for decision-making applications. The foremost challenge is the resilience of clustering models to adversarial noise, which may obscure the result and hinder the models' ability to measure/interpret the data. Unlike supervised learning models and because many clustering algorithms are not differentiable, we have fewer options for applying standard gradient-based adversarial defenses. The literature suggests that little attention has been given to black-box adversarial attacks that exploit clustering vulnerabilities without knowledge of the model [28]. Another growing challenge lies in the ability to manipulate clustering results via adversarial examples mixed with data poisoning attacks. With malicious data input, adversaries may mislead common clustering algorithms such as K-means and Gaussian mixture models with adversarial examples, potentially altering decision boundaries [46]. Beyond model performance hindrances, privacy raises challenges associated with adversarial examples, causing personal data to be effectively unlearnable. Despite these challenges being noted and the problems accepted, the body of research evaluating adversarial threats in an unsupervised learning context remains limited. Further research is necessary on real-world adversarial threats in an unsupervised learning context associated with developing greater resilience to threats and developing robust mechanisms to implement adversarial transferability between unsupervised and supervised models.

5.3 Recommendations and Future Work

This section aims to analytically review the recommendations made by various researchers and the proposed next steps in the field. Figure 9 offers a concise representation of the recommendations made, which includes the researchers' recommendations and insight into the next steps in research. The next subsections detail the proposed next steps for the field.

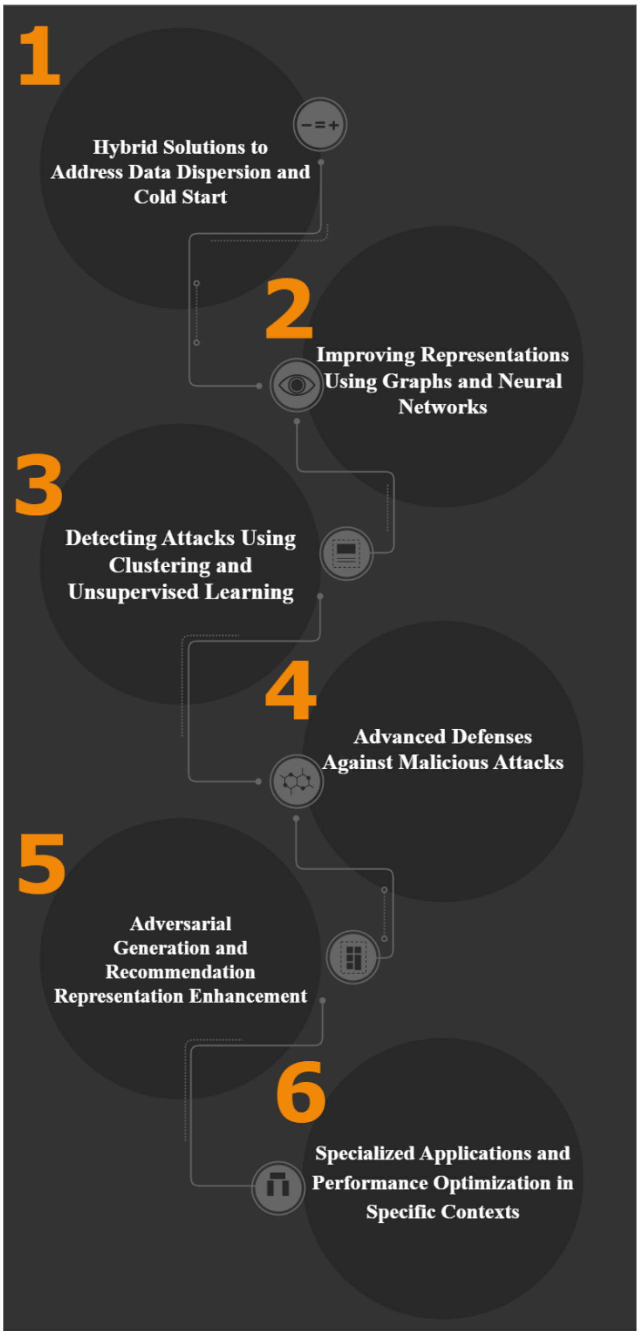
1
Hybrid Solutions to Address Data Dispersion and Cold Start
2
Improving Representations Using Graphs and Neural Networks
3
Detecting Attacks Using Clustering and Unsupervised Learning
4
Advanced Defenses Against Malicious Attacks
5
Adversarial Generation and Recommendation Representation Enhancement
6
Specialized Applications and Performance Optimization in Specific Contexts

Fig. 9 Recommendations and future work

5.3.1 Hybrid Solutions to Address Data Dispersion and Cold Start

Recent research highlights the value of implementing content-based filtering, collaborative filtering, clustering approaches, and deep learning models together to balance multiple recommendation contexts and improve outcomes. One approach is the DAAN model, which uses both domain-shared and domain-specific knowledge to fill in sparse user–item interaction matrices in cross-domain recommendations [66]. The DAAN model uses attention to determine whether it is less relevant to recommend domain-shared knowledge or domain-specific knowledge and suggests relevance in recommendation. In future research, we expect to include things such as adding auxiliary information from user reviews and item content to create better recommendations and expanding their recommendations to other tasks beyond just item recommendations, e.g., predicting ratings. Similar hybrid recommendation algorithms that combine content filtering and CF outperform single methods when dealing with sparse datasets [4]. To this extent, these hyperparameter-tuning approaches create better outcomes in the selection of nearest neighbors and the details of similarity filtering, which are relevant to their adaptation to new users and items. Moreover, clustering-based recommendation systems (CBRS) are being developed to address scalability and diversity and even utilize a vector space model borrowing from information retrieval [39]. Future research and development involving hybrid clustering and consensus clustering should improve CBRS performance. In addition to clustering and filtering, CF approaches using machine learning models have been used to improve personalized recommendations, particularly in domains such as movie recommendations [43]. Co-clustering and slope-one methods could be added to provide even more successful recommendation results. Cross-domain recommendation systems could provide even more improvement by integrating time, trust, context, location and/or sentiment analysis with the aim of enhancing recommendation quality. The Trust-Aware Spatial-Temporal Activity-Based Denoising Autoencoder (TSTDAE) method can minimize cold start issues and provide filtered biased user information while clustering users with similar behavior in context toward a recommendation [19]. Not only could TSTDAE provide context-aware recommendations to users, but it was also able to recommend the best timing to implement the recommendation to maintain user interest and provide user satisfaction. Finally, in developing and overcoming the challenges of sequential recommendations, distributed learning approaches such as DistVAE offer a new way of doing this [32]. The significant aspect of DistVAE is that it functions as a client–server architecture and protects data informatics. Using clustering and the GMM, the DistVAE minimizes gradient noise, mitigates issues of low strategy adoption and builds stability for learning. By using masked attention layers, DistVAE improves the modelling of long-term dependencies, creating a richer user-context environment, which aggregates useful user information in providing greater recommendations. The results of their example show that DistVAE perceives better outcomes than centralized versions of DistVAE do and thus greater usability and relevance to real-world applications. Overall, numerous hybrid approaches indicate that integrating various recommendations may facilitate the handling of their identified goals of sparsity, cold start, and scalability. Future work will likely explore the development of hybrid methods through more advanced clustering modalities, aiding auxiliary data, and creating more adaptive frameworks for learning to streamline the adaptability and effectiveness of RSs.

5.3.2 Improving Representations Using Graphs and Neural Networks

A key starting point for improving RSs is improving representation learning. Many studies are beginning to pay attention to the role of graph representation learning, neural networks, and hybrid methods in developing representations of users and items, enhancing clustering, and improving recommendation quality across various applications. One study discussed travel recommendation system improvements that could be achieved by using variational autoencoders and other generative models to enhance the specifics of existing models (e.g., implementing state-of-the-art models and fine-tuning) and improve performance [68]. They mention composite models that take advantage of contextual information and utilize CF techniques to more accurately refine destination recommendations tailored to individual travellers and develop improved methods to understand complex relationships between POI and user preferences to increase accuracy. An area of study may focus on adversarial learning tasks to improve trip recommendation via graphs as well. Additionally, another study revealed that the AG-cluster also contains off-the-shelf attention mechanisms employing a GCN, which are improvements over previous clustering methods (e.g., item–item similarity), yielding results that are both more accurate and robust when predicting POI stored by users in a personalized recommendation framework [47]. Another study described future directions for exploring alternative network representations rather than just relying on node2vec alone for use in RSs [38]. In addition, a promising area of research would be to explore how combining various recommendation algorithms and utilizing network embedding methods would improve representation learning and recommendation quality/results [39]. In another area, online education, this study explores the use of clustering followed by deep learning processes to enhance course recommendations. The grouped course recommendations describe integrated bidirectional long short-term memory (BiLSTM) and multilayer perceptrons (MLPs) to provide users with scalable and personalized recommendations to maximize learner engagement and results [41]. The continued target of future work will be to enhance the dynamic clustering updates in their project and hybrid recommendations, which incorporate the same approaches using various CF techniques to bring these hybrids into more robust methods. Finally, in a capture of importance based on observed signals in estimating items and entities to potentially update beyond classical/implicit observations if just to enrich recommendations, expanding representations include distinctive signal observations and their integration to enhance description/assessment, leading to rationalized estimation [44]. Including large language models (LLMs) may also capture unique node representations and identify additional collaborative signal relationships, which could direct greater accuracy to recommendations.

5.3.3 Detecting Attacks Using Clustering and Unsupervised Learning

The increasing complexity of adversarial attacks on RSs has necessitated the progression of different clustering-based and unsupervised detection methods to specifically detect and mitigate adversarial attacks. One of the papers introduced the InfoAtk framework, which refines adversarial attacks while optimizing its detection evasion, remaining stealthy. The authors emphasize balancing attack effectiveness with stealth and always contend that they want all manipulation to be unnoticed [59]. To help defend against adversarial attacks and in response to the InfoAtk framework, some authors have suggested various semisupervised learning (SSL) detection methods, such as the DSSD-ImMPL and KC-GCN approaches, which rely on leveraging user relationship graphs as well as meta-learning and communication graphs to streamline the approaches and, in general, increase the detection mean accuracy across RSs [59], [63]. Overall, all the papers emphasized the automatic detection of influential users participating in group shilling attacks on recommendation systems. Another part of the discussion also called attention to adversarial-learned injection attacks and showed how final representations using knowledge graphs and VAEs necessitated the reconstruction of embedding vectors and indicated the need to explore malicious profiles with VAEs [3]. This methodology not only detects attacks generated adversarially but also detects attacks generated heuristically; it performs better than normal detection approaches do. With respect to not only RSs but also NIDS, several ensemble clustering methods, such as ECT-Subspace and ECT-Noise, have enhanced classification performance with respect to obfuscated intrusions in NIDS [53]. Rather than relying on defined class labels, such as traditional methods do, ensemble clustering can utilize a generalized feature mapping transformation while still maintaining predictive competency and robustness toward adversarial attacks. Fully unsupervised, divide-and-conquer clustering methodologies, such as the OPTICS algorithm, constitute a means of detecting shilling attacks, at the risk of not providing any prior knowledge or labelled attack profiles to the inspecting user [72]. Attack behaviors can be divided into standard and obfusgated behavior attacks, but detection still remains intact; however, the computational overhead is reduced. Part of this research highlights the preferred delineation of target item aspects for further refinement of methods to characterize attacks.

5.3.4 Advanced Defences Against Malicious Attacks

The existing body of research suggests the importance of adaptive detection modules, hardened training methods, and modelling adversarial behavior to resist these types of attack scenarios [61]. The possibilities for defenses in this domain are vast, but the initial proposed work includes an adaptive fraudster detection module against node injection attacks with a system of TCD for adversarial robustness [60]. Future work will include tests of defenses that are cross-domain and improved attack/defense dynamics via GCoAttack. In CF environments, understanding adversarial strategies and vulnerabilities is essential for strengthening defense mechanisms, particularly against botnets and corrupted benign data. Adversarial training techniques such as Infmix and APT are being explored to enhance security, especially in non-matrix factorization models [62]. Robust training strategies, including stagewise hints training and noise layers, have shown promise in reducing malicious user attacks while maintaining prediction accuracy [65]. Future work will focus on improving anomaly detection, refining clustering techniques, and optimizing poisoning data generation algorithms [31]. Overall, multilayered defense frameworks that integrate fraud detection, adversarial modelling, and cooperative defense strategies are essential for ensuring the long-term security and reliability of RSs in real-world applications.

5.3.5 Adversarial Generation and Recommendation Representation Enhancement

Recent research highlights adversarial learning, GAN-based models, trust management, and normalization techniques as key strategies for enhancing recommendation system resilience and representation learning. One approach leverages cross-domain user profiling to enhance adversarial attacks on black-box RSs, improving attack effectiveness while ensuring targeted item selection [25]. Future work aims to replace existing models with GAN-based embedding frameworks to reduce training costs and improve scalability [70]. Beyond adversarial attacks, trust management frameworks are being explored through GAN-based models, Q-Learning, Federated Learning, and blockchain integration, which enhance training efficiency and security [29]. Additionally, adversarial learning is being applied to representation mechanisms rather than just embeddings, utilizing multi-channel GCN and attention mechanisms to strengthen representation robustness [58]. Adversarial training and normalization techniques are essential for model generalization and stability, with research showing a significant drop in performance when these components are removed [27]. Finally, the GI-AAE demonstrates a group influence-based deep adversarial autoencoder, which improves the recall performance and decision-making quality [42]. To conclude, these studies not only exemplified the necessity of scalable, trust-aware, and adversarially robust recommendation systems but also approached the continuing accuracy and resiliency of future systems.

5.3.6 Specialized Applications and Performance Optimization in Specific Contexts

In recent studies of recommendation systems, One of the studies suggested that knowledge graph modifications, disentangled latent factors, deep learning improvements and clustering methods as the main strategies for enhancing robustness, accuracy and interpretability [54]. Studies mention knowledge graph (KG) modifications to sustain better resilience against poisoning attacks and study hyperactive user behaviour influenced by political agenda-setting in OSNs

[67], [69]. Disentangled latent factors for more interpretable user representations and the use of spectral clustering help to improve the accuracy of cluster estimations with GANs [50]. For news RSs, the combination of graph neural networks, bat optimization, and attention improves semantic vectorization and the relevance of content to the user [37]. In movie recommendation, transformer models improve sequential pattern recognition, increasing recall and accuracy [40]. These findings emphasize the importance of context-specific optimizations in recommendation systems, with future research focusing on clustering techniques, KG adaptations, and deep learning advancements to enhance performance and resilience across diverse domains.

6. A CRITICAL ANALYSIS OF LITERATURE

In this section, we critically examine and analyse the data extracted from the collected research. Table I provides a comprehensive overview of the datasets used, categorizing them based on type, classification, quantity, volume, and domain (e.g., commercial, entertainment, etc.) within the context of RSs, adversarial learning, and clustering. A thorough analysis of these datasets reveals significant trends, strengths, and limitations in existing research methodologies. The following subsections explore these aspects in detail, highlighting key patterns, potential biases, and areas requiring further investigation.

6.1 Prevalence and Standardization of Dataset Usage

One highlighted trend from the literature is the reliance on benchmark datasets such as MovieLens, Amazon, Netflix, and Yelp, which are tremendously popular datasets in recommender system research. Researchers have relied on these data since they are structured, big data, and widely used in the literature. The ubiquity of their usage enables comparability, allowing researchers to benchmark their models against something relatively stable and widely accepted. The reliance on a limited number of state-of-the-art local benchmarks raises concerns about overfitting the methodological approach, considering that any research findings may not be generalizable due to limited application areas. Figure 10 depicts the influence of film-related datasets, which were the most employed datasets in this review. In contrast, social media datasets were incorporated in a highly selective manner, indicating a narrower focus on user interactions within social platforms. Moreover, datasets such as CCV and Yelp occupied an intermediate position, suggesting a moderate level of adoption in studies exploring RSs, clustering, and adversarial learning. This distribution highlights a potential imbalance in dataset selection, with an apparent emphasis on entertainment-based data.

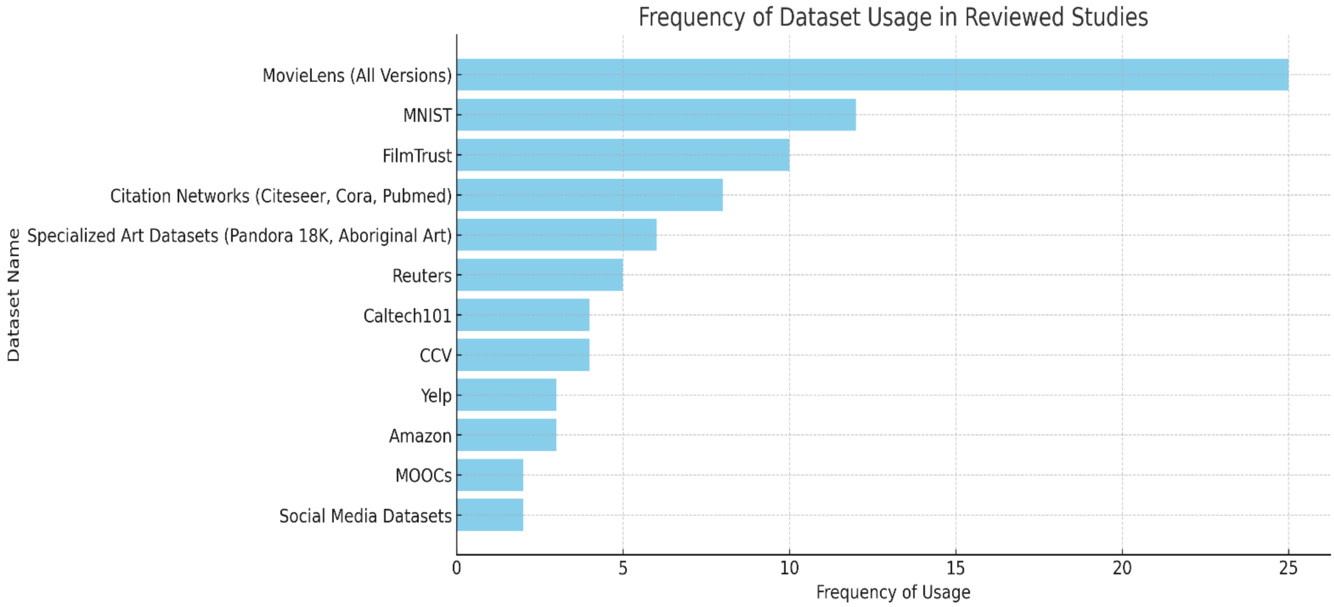

Fig. 10 Frequently used dataset for the SLR

6.2 Nature of Data: Text, Images, Reviews, and Citation Networks

Most studies focus on structured numerical ratings (e.g., MovieLens, Yelp, Netflix) and textual reviews (Amazon, IMDb, Yelp). Image-based datasets (e.g., CIFAR, Market-1501) remain underutilized despite their potential for visual-based recommendations. Similarly, citation networks (Cora, PubMed, Citeseer) appear in clustering and adversarial learning research but are rarely applied to RSs. This suggests an opportunity for adversarial learning to enhance research paper recommendations, fake citation detection, and scholarly influence assessments.

6.3 Redundancy and Dataset Selection Bias

The literature reveals redundancy in dataset selection, with repeated use of benchmark datasets. While standard datasets facilitate reproducibility, they fail to capture the complexities of real-world recommendation environments. The lack of diverse datasets hinders the evaluation of adversarial robustness. Future work should explore real-time streaming data, user behavior logs, and multimodal datasets to enhance recommendation system adaptability.

6.4 Adversarial Threats and Defenses

Adversarial learning has been widely investigated in RSs, clustering, and security applications, yet its full potential remains unexplored. Many studies rely on synthetic attack datasets instead of real-world adversarial cases, limiting their practical applicability. Future research should focus on dynamic social networks, misinformation filtering, financial fraud prevention, and online security systems where adversarial threats evolve over time.

6.5 Neglected Domains: Absence of Financial Datasets and the Need for Adversarial Robustness

In Table I, which extracts all the details and features of the dataset in SLR, a critical gap in the literature is the underrepresentation of financial, healthcare, and other datasets in recommender system research. While recommendation models have been widely applied in entertainment, retail, and e-commerce, their integration into financial domains, such as investment strategies, credit scoring, fraud detection, and risk assessment, remains insufficiently explored. Given the global dependence on financial systems, the absence of adversarially robust recommendation models in this domain presents a significant research shortcoming. Financial recommendation systems are highly susceptible to adversarial attacks, which can manipulate investment recommendations, exploit trading algorithms, or deceive credit scoring models. The lack of adversarial learning strategies in financial RSs leaves them vulnerable to data poisoning, model inversion, and recommendation bias, potentially leading to severe financial fraud, misinformation, and market manipulation. Future research should prioritize the integration of adversarial defenses into financial recommendation models to enhance their resilience and robustness against such threats. This includes developing robust fraud detection mechanisms, ensuring fairness in loan approval RSs, and securing financial trading algorithms from adversarial exploitation. Expanding the dataset diversity beyond traditional entertainment-based benchmarks will also improve the applicability of recommendation systems in high-stakes financial decision-making.

This study proposes the utilization of financial data, specifically the "All Lending Club loan data", which is a comprehensive financial dataset designed to track loan performance. The Lending Club dataset is characterized by its high-dimensional nature, comprising 151 distinct features, making it a rich source of information for financial analysis. Additionally, the dataset qualifies as a "large dataset" with more than 1,048,576 samples and a total size exceeding 3.2 gigabytes, which presents both challenges and opportunities for data processing and analysis. Despite the substantial volume and dimensionality of the dataset, it remains neither complex nor challenging for application to the three key concepts introduced in this study. Additionally, concepts of parameter flexibility, robustness, and scalability can be effectively implemented, ensuring that the dataset can be leveraged for accurate and reliable results. The flexibility of the dataset allows for various analytical approaches, the robustness ensures that the findings are resilient under different conditions, and the scalability facilitates the processing of large datasets, thereby enabling the extraction of meaningful insights. Consequently, this study highlights the potential of "All Lending Club loan data" as a tool for applying and testing the proposed methodologies in financial data analysis.

TABLE I Description dataset used and information extraction from the SLR.

Ref	Dataset	Dataset resource/availability	Sample	NOF	Dataset type
[48]	1- CCV 2- MSRC-V1 3- Reuters 4- MINIST 5- CaLTECH101-20 6- VOC	1- https://www.ee.columbia.edu/ln/dvmm/CCV/ 2- https://linqs-data.soe.ucsc.edu/public/lbc/ 3- https://kdd.ics.uci.edu/databases/reuters21578/ 4- http://yann.lecun.com/exdb/mnist/ 5- http://www.vision.caltech.edu/Image-Datasets/Caltech101/ 6- https://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2008/	1- 6773 2- 210 3- 1200 4- 2000 5- 2386 6- 5649	1- N/A 2- N/A 3-N/A 4- 784 5- N/A 6- N/A	Image + text
[45]	1- UCI Handwritten Digits 2- MNIST 3- MoCap Hand Postures	1- https://archive.ics.uci.edu/dataset/236/seeds 2- https://www.kaggle.com/datasets/hojjatk/mnist-dataset 3- https://archive.ics.uci.edu/dataset/391/mocap+hand+postures	1-70 2-200 3-200	N/A	Image
[51]	ASNM dataset	https://ieeexplore.ieee.org/document/9115004/	11,445	14	Text
[47]	1-DBLP 2-ACM 3-Citeseer 4-Cora 5-Pubmed 6-Karate 7-Game of Thrones	1- https://www.kaggle.com/datasets/dheerajmpai/dblp2023 2- https://paperswithcode.com/dataset/acm 3- https://paperswithcode.com/dataset/citeseer 4- https://paperswithcode.com/dataset/cora 5- https://paperswithcode.com/dataset/pubmed 6- https://networkrepository.com/soc-karate.php 7- N/A	1- N/A 2- N/A 3- nodes =3327, edges =4732 4- nodes =2708, edges =5429 5- nodes=19717, edges =44338 6- N/A 7- Nodes= 107, edges=353	1-N/A 2- N/A 3- 3703 4- 1433 5- 500 6-N/A 7-N/A	Citation network
[57]	MNIST	https://www.kaggle.com/datasets/hojjatk/mnist-dataset	handwritten digits (0–9)	4	Image
[49]	1- MNIST 2- Fashion-MNIST 3- 10x_73k 4- Pendigit	1- https://www.kaggle.com/datasets/hojjatk/mnist-dataset 2- https://www.kaggle.com/datasets/zalando-research/fashionmnist 3- N/A 4- https://archive.ics.uci.edu/dataset/81/pen+based+recognition+of+handwritten+digits	Sampling (Data Dimension) 1- 70,000 (28 × 28) 2- 70,000(28 × 28) 3- 73,233(1 × 720) 4- 10,992(1 × 16)	N/A	1- Image 2- Image 3- RNA discrete data 4- Time sequences
[54]	1- early-Twitter 2- late-Twitter 3- Synthetic	1- N/A 2- N/A 3- N/A	1- 4312 2- 4312 3- 4000	N/A	Network structures + text
[52]	1- DS-1 2- DS-2 3- DS-3 4- DS-4	1 & 2 = https://phishtank.com/ http://www.Alexa.com 3-https://archive.ics.uci.edu/ 4- https://data.mendeley.com/	Sampling (Leg%, phi%)[1] 1. 2210 (44.7,55.2) 2. 11055 (55.6,44.3) 3. 1250 (43.8,56.1) 4. 10000 (50,50)	1. 7 2. 30 3. 9 4. 48	Text

Ref	Dataset	Dataset resource/availability	Sample	NOF	Dataset type
[53]	ASNM dataset	https://ieeexplore.ieee.org/document/9115004/	11,445	176	Text
[55]	1. Market-1501 2. DukeMTMC-reID 3. MSMT17	1. https://paperswithcode.com/dataset/market-1501 2. https://paperswithcode.com/dataset/dukemtmc-reid 3. https://paperswithcode.com/dataset/msmt17	1. 36,036 2. 36,411 3. 126,441	2048	Image
[50]	1. Coil20 2. Extended YaleB 3. Orl 4. Coil100 5. USPS 6. MNIS	1. https://www.kaggle.com/datasets/cyx6666/coil20 2. https://paperswithcode.com/dataset/extended-yale-b-1 3. https://paperswithcode.com/dataset/orl 4. https://www.kaggle.com/datasets/jessicali9530/coil100 5. https://paperswithcode.com/dataset/usps 6. https://www.kaggle.com/datasets/hojjatk/mnist-dataset	1. 1,440 2. 2,432 3. 400 4. 7,200 5. 9,298 6. 70,000	1. 1,024 2. 1,024 3. 4,096 4. 1,024 5. 256 6. 784	Image
[58]	1. Cornell 2. Texas 3. Washington 4. Wisconsin 5. Citeseer 6. Cora 7. Pubmed	1,2,3,4:https://paperswithcode.com/dataset/webkb 5. https://paperswithcode.com/dataset/citeseer 6. https://paperswithcode.com/dataset/cora 7. https://paperswithcode.com/dataset/cora	No Nodes Edges 1 195 304 2 183 328 3 217 446 4 262 530 5 3,312 4,732 6 2,708 5,429 7 19,717 44,338	1- 1703 2- 1703 3- 1703 4- 1703 5- 3703 6- 1433 7- 500	Citation Networks
[56]	1. Dataset 1 2. Dataset 2 3. Dataset 3	N/A	1. 4,105 2. 18,038 3. 5,313	N/A	Images
[21]	1. PU 2. CWRU 3. IMS 4. XJTU-SY 5. wheelset	1. https://mb.uni-paderborn.de/kat/forschung/kat-datacenter/bearing-datacenter/data-sets-and-download 2. https://www.kaggle.com/datasets/brjapon/cwru-bearing-datasets 3. https://paperswithcode.com/dataset/ims-bearing-dataset 4. https://www.kaggle.com/datasets/zwming/xjtu-sy 5. https://www.kaggle.com/datasets/sravanchittupalli/wheels-dataset	1. 24,000 2. 13,200 3. 3,600 4. 3,600 5. 8,400	All= 1200	Time-Series
[28]	1. Fashion-MNIST 2. CIFAR-10 3. 20-Newsgrous 4. UCI Digits	1. https://www.kaggle.com/datasets/zalando-research/fashionmnist 2. https://paperswithcode.com/dataset/cifar-10 3. https://www.kaggle.com/datasets/crawford/20-newsgroups 4. https://archive.ics.uci.edu/dataset/80/optical+recognition+of+handwritten+digits	NO Size Subset 1 70,000 1,600 2 60,000 1,600 3 20,000 1,600 4 5,620 N/A	1- 784 2- 2,048 3- 80 4- 64	Image

Ref	Dataset	Dataset resource/availability	Sample	NOF	Dataset type
[46]	1. Iris 2. MNIST 3. Fashion-MNIST 4. CIFAR-2 5. TIMIT	1. https://www.kaggle.com/datasets/vikrishnan/iris-dataset 2. https://www.kaggle.com/datasets/hojjatk/mnist-dataset 3. https://www.kaggle.com/datasets/zalando-research/fashionmnist 4. https://paperswithcode.com/dataset/cifar-10 5. https://paperswithcode.com/dataset/timit	1. 150 2. 70,000 3. 60,000 4. 10,000 5. 630	1- 4 2- 784 3- 784 4- 3,072 5- 39	1. Text 2. Image 3. Image 4. Image 5. audio
[27]	1- Citeseer 2- Cora 3- Pubmed	1- https://paperswithcode.com/dataset/acm 2- https://paperswithcode.com/dataset/citeseer 3- https://paperswithcode.com/dataset/cora	NO Node Edges 1 3,327 4,732 2 2,708 5,429 3 19,717 44,338	1-3703 2-1433 3-500	Citation Networks
[23]	1. YelpCHI 2. Movies	1. https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset 2. https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows	Users Items Edges Fake Users 38,063 201 67,395 7,739 39,578 71,187 232,082 19,909	N/A	Text (reviews) + numerical (ratings).
[59]	1. ML-100k 2. ML-1M 3. Douban 4. Epinions	1. https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset 2. https://www.kaggle.com/datasets/prajitdatta/movielens-100k-dataset 3. https://www.kaggle.com/datasets/fengzhujoey/douban-datasetratingreviewside-information 4. https://paperswithcode.com/dataset/epinion	Users Items [2]Int [3]Spa 943 1,682 100,000 93.6% 6,040 3,952 1,000,000 95.8% 2,831 36,821 805,611 99.2% 46,846 40,706 305,249 99.9%	N/A	Text (reviews) + numerical (ratings).
[60]	1. FilmTrust 2. ML-100k 3. ML-1M	1. https://www.kaggle.com/datasets/abdelhakaissat/film-trust 2. https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset 3. https://www.kaggle.com/datasets/prajitdatta/movielens-100k-dataset	Users Items Ratings [3]Spa 796 2011 30880 98.07 943 1682 100,000 98 6040 3706 1000,209 95.5	1- 128 2- 128	Text (reviews) + numerical (ratings
[63]	1. MovieLens 10M 2. FilmTrust 3. Amazon	1. https://grouplens.org/datasets/movielens/10m/ 2. https://www.kaggle.com/datasets/abdelhakaissat/film-trust 3. https://www.amazon.com/	Users Items Ratings 71,567 10,681 10,000,054 1,227 2,059 34,886 [4](GE=2,275, ATT=1,508) 16,885 47,408	N/A	1. Text (reviews) + numerical (ratings 2. Text (reviews) + numerical (ratings 3. User-item interactions with labelled attack/genuine users
[61]	1. Jester 2. Movie 3. E-Shopping	1. https://grouplens.org/datasets/jester/ 2. N/A 3. N/A	Users Items Ratings 3,000 150 ~30,000 10,000 23,000 181,382 5,000 40,000 420,292	N/A	Text (reviews) + numerical (ratings

Ref	Dataset	Dataset resource/availability	Sample	NOF	Dataset type
[64]	1. Netflix 2. Amazon	1. https://paperswithcode.com/dataset/netflix-prize 2. https://www.amazon.com/	Users · Items · Ratings 2,000 · 4,000 · 215,884 5,055 · 17,610 · 53,777	1. 4 2. 4	Text (reviews) + numerical (ratings
[62]	1. FilmTrust 2. ML-100k 3. ML-1M 4. Yelp	1. https://www.kaggle.com/datasets/abdelhakaissat/film-trust 2. https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset 3. https://www.kaggle.com/datasets/prajitdatta/movielens-100k-dataset 4. https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset	Users · Items · ratings · ^{3}Spa 796 · 2011 · 30880 · 98.07 943 · 1682 · 100,000 · 93.7 6040 · 3706 · 1000,209 · 95.5 14575 · 25602 · 569949 · 99.8	N/A	Text (reviews) + numerical (ratings)
[68]	1. Edinburgh 2. Glasgow 3. Osaka 4. Toronto	https://www.kaggle.com/datasets/javidtheimmortal/yfcc100msfmdataset	Users · POI Visits · Trips 1,454 · 33,944 · 5,028 601 · 11,434 · 2,227 450 · 7,747 · 1,115 1,395 · 39,419 · 6,057	N/A	Trajectory dataset
[65]	1. ML-100k 2. ML-1M	1. https://www.kaggle.com/datasets/prajitdatta/movielens-100k-dataset 2. https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset	Users · Items · ratings 943 · 1682 · 100,000 6040 · 3952 · 1000,209	N/A	Text (reviews) + numerical (ratings).
[66]	1. Amazon 2. Netflix & MovieLens	1. http://jmcauley.ucsd.edu/data/amazon/ 2. https://grouplens.org/datasets/movielens/ 3. https://www.kaggle.com/netflix-Inc./netflix-prize-data	Task · Users · Items · Ratings · Spa Task 1 · 1,640 · 18,025 · 72,195 · 99.76% - · 1,640 · 5,295 · 24,186 · 99.72% Task 2 · 835 · 5,852 · 15,499 · 99.68% - · 835 · 5,976 · 15,055 · 99.70% Task 3 · 3,820 · 46,318 · 321,649 · 99.82% - · 3,820 · 7,641 · 61,222 · 99.79%	N/A	Text (reviews) + numerical (ratings).
[25]	1. ML10M 2. ML20M	1. https://www.kaggle.com/datasets/amirmotefaker/movielens-10m-dataset-latest-version 2. https://www.kaggle.com/datasets/grouplens/movielens-20m-dataset	Domain · Users · Items · Int Target (ML10M) · 19,267 · 6,984 · 437,746 Source (Flixster) · 93,702 · N/A · 4,680,700 Target (ML20M) · 38,087 · 8,325 · 838,491 Source (Netflix) · 474,471 · N/A · 62,937,958	1. 8 2. 8	Text (reviews) + numerical (ratings).
	1. Amazon 2. MovieLens-basic 3. MovieLens-adv	1. https://www.amazon.com/ 2. https://grouplens.org/datasets/movielens/100k/ 3. N/A	Leg · ^{5}MaU · Items · Ratings 3,118 · 1,937 · 16,885 · 51,346 943 · 168 · 1,682 · 127,416 943 · 169 · 1,682 · 123,407	1. 32 2. 32 3. 32	Malicious user detection

<table>
<tr><th>Ref</th><th>Dataset</th><th>Dataset resource/availability</th><th>Sample</th><th>NOF</th><th>Dataset type</th></tr>
<tr><td>[67]</td><td>1. MovieLens-1M
2. Fund</td><td>1. https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset
2. N/A</td><td><table><tr><th>Users</th><th>Items</th><th>Int</th><th>[6]KGT</th></tr><tr><td>6,036</td><td>2,347</td><td>753,772</td><td>20,195</td></tr><tr><td>90,218</td><td>2,368</td><td>698,140</td><td>6,312</td></tr></table></td><td>1. 64
2. 64</td><td>knowledge graph poisoning</td></tr>
<tr><td>[3]</td><td>1. DS1=MovieLens-1M
2. DS2=Book-Crossing
3. DS3=Nowplaying</td><td>1. https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset
2. https://www.kaggle.com/datasets/somnambwl/bookcrossing-dataset
3. https://www.kaggle.com/datasets/chelseapower/nowplayingrs</td><td><table><tr><th>Metric</th><th>DS1</th><th>DS2</th><th>DS3</th></tr><tr><td>Domain</td><td>Movies</td><td>Books</td><td>Music</td></tr><tr><td>Ratings</td><td>753,772</td><td>69,872</td><td>361,346</td></tr><tr><td>Users</td><td>6,036</td><td>17,860</td><td>4,776</td></tr><tr><td>Items</td><td>2,445</td><td>14,967</td><td>26,911</td></tr><tr><td>KG Entities</td><td>182,011</td><td>77,903</td><td>Valence-based links</td></tr><tr><td>KG Relations</td><td>12</td><td>25</td><td>N/A</td></tr></table></td><td>1. 261
2. 261
3. 261</td><td>knowledge graph</td></tr>
<tr><td>[2]</td><td>1. Movielens-10M
2. Amazon</td><td>1. https://www.kaggle.com/datasets/amirmotefaker/movielens-10m-dataset-latest-version
2. https://www.amazon.com/</td><td><table><tr><th>Metric</th><th>Movielens-10M</th><th>Amazon</th></tr><tr><td>Domain</td><td>Movies</td><td>Products</td></tr><tr><td>Total Users</td><td>71,567</td><td>4,902</td></tr><tr><td>Total Items</td><td>10,681</td><td>16,885</td></tr><tr><td>Ratings</td><td>10,000,054</td><td>51,346</td></tr><tr><td>Attack Types</td><td>Random, Average, Bandwagon, GSA-GAN</td><td>Real-world attacks</td></tr></table></td><td>N/A</td><td>Text (reviews) + numerical (ratings).</td></tr>
<tr><td>[70]</td><td>1. Ciao
2. LastFM</td><td>1. https://www.kaggle.com/datasets/aravindaraman/ciao-data
2. files.grouplens.org/datasets/hetrec2011/</td><td><table><tr><th>Users</th><th>Items</th><th>Int</th><th>Domain</th></tr><tr><td>996</td><td>1,927</td><td>18,648</td><td>E-commerce</td></tr><tr><td>1,892</td><td>17,631</td><td>92,834</td><td>Music</td></tr></table></td><td>N/A</td><td>Text (reviews) + numerical (ratings).</td></tr>
<tr><td>[69].</td><td>Political Facebook Pages</td><td>N/A</td><td>1. 4 M users, 3 M+ reactions</td><td>4</td><td>Interaction-based</td></tr>
<tr><td>[5]</td><td>1. MovieLens-1M
2. CiteULike
3. LastFM</td><td>1. https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset
2. https://paperswithcode.com/sota/recommendation-systems-on-citeulike
3. files.grouplens.org/datasets/hetrec2011/</td><td><table><tr><th>Users</th><th>Items</th><th>Int</th><th>Spa</th></tr><tr><td>6,040</td><td>3,544</td><td>993,482</td><td>95.4%</td></tr><tr><td>5,551</td><td>16,980</td><td>204,986</td><td>99.8%</td></tr><tr><td>1,892</td><td>17,632</td><td>92,834</td><td>97.3%</td></tr></table></td><td>N/A</td><td>Text + ratings</td></tr>
<tr><td>[37]</td><td>1. DS1=Baidu News
2. DS2=MIND
3. DS3=Adressa
4. DS4=Digg</td><td>1. https://www.kaggle.com/code/mpwolke/baidu-news
2. https://www.kaggle.com/datasets/arashnic/mind-news-dataset
3. https://www.kaggle.com/datasets/ayushhirdani/adressa
4. N/A</td><td><table><tr><th>Users</th><th>Items</th><th>Int</th></tr><tr><td>853</td><td>N/A</td><td>3,654</td></tr><tr><td>920,056</td><td>N/A</td><td>6,325</td></tr><tr><td>3,083,438</td><td>48,486</td><td>27,223,576</td></tr><tr><td>139,409</td><td>3,553</td><td>3,018,197</td></tr></table></td><td>N/A</td><td>1. Text
2. Text
3. User interactions
4. Social links</td></tr>
</table>

Ref	Dataset	Dataset resource/availability	Sample	NOF	Dataset type
[38]	1. Yelp (Pittsburgh) 2. Yelp (Madison) 3. Amazon 4. MovieLens	1. https://www.kaggle.com/datasets/mobasshir/yelpdata 2. https://www.kaggle.com/datasets/mobasshir/yelpdata 3. https://www.amazon.com/ 4. N/A	User Item Link category 466 1672 10373 161 332 1172 5597 150 6831 32054 71661 912 943 1682 100000 50	1- 100 2- 100 3- 100 4- 100	Business Reviews Business Reviews Product Reviews Movie Ratings
[1]	1. MovieLens 100K 2. MovieTweetings 10K 3. FilmTrust	1. https://www.kaggle.com/datasets/prajitdatta/movielens-100k-dataset 2. https://www.kaggle.com/datasets/tunguz/movietweetings 3. https://www.kaggle.com/datasets/abdelhakaissat/film-trust	Users Items Ratings 943 1,682 100,000 123 3,096 2,233 1,508 2,071 35,497	N/A	Text (reviews) + numerical (ratings)
[4]	1. MovieLens 2. Scientific-Literature	1. https://grouplens.org/datasets/movielens/ 2. N/A	Users Items Ratings 248 1,120 12,500 N/A N/A N/A	N/A	1. Rating 2. Text
[39]	1. LDOS-CoMoDa 2. InCarMusic 3. Apps in Frappe' 4. POI in STS 5. Hotels in -TripAdvisor 6. Apple Store 7. Drug Review	1. https://www.lucami.org/en/research/ldos-comoda-dataset/ 2. N/A 3. https://huggingface.co/datasets/abadesalex/Frappe-mobile-app-usage 4. N/A 5. https://www.kaggle.com/datasets/andrewmvd/trip-advisor-hotel-reviews 6. https://www.kaggle.com/datasets/gauthamp10/apple-appstore-apps 7. https://archive.ics.uci.edu/dataset/461/drug+review+dataset+druglib+com	Domain Users Items Ratings Movies 189 3,029 4,316 Songs 43 139 4,012 Mobile app usage 1,000 ~24,000 ~24,000 Points of Interest 239 184 1,379 Hotel reviews 2,731 2,269 14,175 iOS app ratings 7,197 7,000+ N/A Patient drug reviews 4,143 645 N/A	1- N/A 2- 8 3- N/A 4- N/A 5- N/A 6- N/A 7- N/A	Structured ratings
[43]	1. MovieLens 25M	https://grouplens.org/datasets/movielens/25m/	1. 25 million ratings (used 697,561) 2. 1 million tags 3. 62,000 movies (62,423 used) 4. 162,000 users 5. 99% Sparsity	N/A	Structured ratings
[19]	1. AliExpress	https://www.kaggle.com/datasets/abdullahbuzaid/ali-express-data	Aspect Details Domains Source: Hair & Wigs, Home Appliances. Target: Apparel Accessories, Education & Office Supplies. Size ~2,000–4,500 users; ~1,500–2,900 items; ~2,000–4,600 ratings per domain.	100	E-commerce
[40]	1. Movielens	https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset	1. 3,900 movies 2. 1,000,209 ratings 3. 6,040 users	N/A	Structured ratings

<table>
<tr><th>Ref</th><th>Dataset</th><th>Dataset resource/availability</th><th>Sample</th><th>NOF</th><th>Dataset type</th></tr>
<tr><td>[41]</td><td>MOOCs</td><td>https://www.kaggle.com/discussions/general/307061</td><td>12,340</td><td>5</td><td>Text</td></tr>
<tr><td>[42]</td><td>1. MovieLens 100K
2. MovieLens-1M
3. FilmTrust</td><td>1. https://www.kaggle.com/datasets/prajitdatta/movielens-100k-dataset
2. https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset
3. https://www.kaggle.com/datasets/abdelhakaissat/film-trust</td><td><table>
<tr><th>Ratings</th><th>Users</th><th>Items</th><th>Spa</th></tr>
<tr><td>100,000</td><td>943</td><td>1,682</td><td>93.70%</td></tr>
<tr><td>1,000,209</td><td>6,040</td><td>3,952</td><td>95.74%</td></tr>
<tr><td>35,497</td><td>1,508</td><td>2,701</td><td>98.90%</td></tr>
</table></td><td>N/A</td><td>Text + ratings</td></tr>
<tr><td>[44]</td><td>1. DS1=MovieLens-1M
2. DS2=Amazon-Book
3. DS3=LastFM</td><td>1. https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset
2. https://www.kaggle.com/datasets/bittupanchal/amazon-books-dataset
3. files.grouplens.org/datasets/hetrec2011/</td><td><table>
<tr><th>Metric</th><th>DS1</th><th>DS2</th><th>DS3</th></tr>
<tr><td>Users</td><td>1,872</td><td>70,769</td><td>6,036</td></tr>
<tr><td>Items</td><td>3,846</td><td>24,915</td><td>2,478</td></tr>
<tr><td>Interactions</td><td>21,173</td><td>652,614</td><td>306,937</td></tr>
<tr><td>KG Entities</td><td>9,366</td><td>29,713</td><td>102,569</td></tr>
<tr><td>KG-Relations</td><td>60</td><td>39</td><td>32</td></tr>
<tr><td>KG Triplets</td><td>15,518</td><td>686,514</td><td>499,474</td></tr>
<tr><td>Embedding-Size</td><td>64</td><td>64</td><td>64</td></tr>
</table></td><td>64</td><td>Knowledge Graph</td></tr>
<tr><td>[71]</td><td>1. DS1=Synthetic
2. DS2=Movieles
3. DS3=Amazon
4. DS4=Yelp
5. DS5=LastFM</td><td>1. N/A</td><td><table>
<tr><th>Dataset</th><th>Original Size</th><th>Subset</th></tr>
<tr><td>DS1</td><td>1,000 users, 1,000 arms</td><td>-</td></tr>
<tr><td>DS2</td><td>2,113 users, 10,197 movies</td><td>1,000 users, 1,000 items</td></tr>
<tr><td>DS3</td><td>1,429 users, 900 items</td><td>1,400 users, 800 items</td></tr>
<tr><td>DS4</td><td>1,987,929 users, 150,346 items</td><td>2,000 users, 2,000 items</td></tr>
<tr><td>DS5</td><td>1,892 users, 17,632 artists</td><td>500 users, 2,000 arms</td></tr>
</table></td><td>50</td><td>User behavior in recommendation systems</td></tr>
<tr><td>[32]</td><td>1. DS1=ML-latest
2. DS2=ML-1M
3. DS3=MovieTweetings
4. DS4=PEEK</td><td>1. https://grouplens.org/datasets/movielens/latest/
2. https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset
3. https://www.kaggle.com/datasets/tunguz/movietweetings
4. https://github.com/sahanbull/PEEKC-Dataset</td><td><table>
<tr><th>Dataset</th><th>Records</th><th>Users</th><th>Items</th><th>^{7}Avg</th><th>Spa</th></tr>
<tr><td>DS1</td><td>46K</td><td>604</td><td>7,363</td><td>76.2</td><td>98.9%</td></tr>
<tr><td>DS2</td><td>580K</td><td>6,034</td><td>3,533</td><td>95.3</td><td>97.3%</td></tr>
<tr><td>DS3</td><td>57K</td><td>3,758</td><td>7,418</td><td>15.2</td><td>99.8%</td></tr>
<tr><td>DS4</td><td>56K</td><td>7,152</td><td>7,031</td><td>7.8</td><td>99.8%</td></tr>
</table></td><td>1- 128
2- 128
3- 128
4- 128</td><td>1. Movie Recommendations
2. Movie Recommendations
3. Movie Ratings
4. Educational Videos</td></tr>
<tr><td>[31]</td><td>1. Beauty
2. Sports</td><td>1. https://www.kaggle.com/datasets/satrapankti/amazon-beauty-product-recommendation
2. https://www.kaggle.com/datasets/deovcs/amazon-dataset</td><td><table>
<tr><th>Metric</th><th>DS1</th><th>DS2</th></tr>
<tr><td>Users</td><td>22,363</td><td>35,598</td></tr>
<tr><td>Items</td><td>12101</td><td>18357</td></tr>
<tr><td>Int</td><td>198502</td><td>296337</td></tr>
<tr><td>Fake Users</td><td>1</td><td>1</td></tr>
<tr><td>Proportion</td><td>0.0045%</td><td>0.0028%</td></tr>
</table></td><td>N/A</td><td>1. Beauty Products
2. Amazon Sports</td></tr>
</table>

Ref	Dataset	Dataset resource/availability	Sample			NOF	Dataset type
			Metric	DS1	DS2		
[72]	1. DS1=MovieLens-100K 2. DS2=Netflix	1. https://www.kaggle.com/datasets/prajitdatta/movielens-100k-dataset 2. https://www.kaggle.com/datasets/victorsoeiro/netflix-tv-shows-and-movies	Users	943	2,000	N/A	1. Movie Recommendations 2. Movie Recommendations
			Items	1,682	4,000		
			Ratings	100,000	280,015		
			Attack Focus	Shilling	Shilling		

(NOF= Number of features), [1] (Leg= Legitimate, Phi= Phishing), [2] (Int=Interaction), [3] (Spa=Sparsity), [4] (GE=genuine, ATT=attack), [5] (MaU=Malicious Users), **[6] (**KGT **=**KG Triples), (DS= Dataset), [7] (Avg =Avg. Sequence Length)

6.6 Evaluation and Comparative Analysis

As previously presented, this study is based on a classification based on four axes, which are the intersection between the three concepts, in addition to the integration among the three concepts combined with each other. Table 2 shows the extracted results for each study from the literature, where the first column represents the reference, the second column represents the data used, and the third column categorizes the work by the methods employed in each study, while the remaining part of this column presents the metrics for evaluating the results of these methods. in the literature is based on specific criteria based on its specific axis. Each author used methods based on what they deemed appropriate to achieve high results and accuracy. For example, [48] presents the results when three datasets are used and the method AGARL is applied. In addition, four evaluation criteria were adopted: accuracy (ACC), normalized mutual information (NMI), precision (PRE), and purity (PUR). The proposed method was compared with other methods from the literature, and based on the results indicated by the author, the proposed model achieved the highest results on all announced criteria. In addition, owing to the use of more than one dataset, this work demonstrated that the proposed method can be generalized because three datasets were used for the proposed method and achieved high results and accuracy. While [37] presented a proposed method that was compared with four other methods used in the literature, this proposed method reflected the accuracy described by the authors and the flexibility of implementation. This work was applied to four news datasets (Chinese, English, Addressa, and Digg). The quality and originality of the models were evaluated based on three criteria: PRE, recall (REC), and F1-score (F1). The proposed model, as explained by the authors, achieved better results than the methods used in the literature did, leading to a widespread perception that the model proposed by them provides high performance. In contrast, the fourth axis, which is built on the integration of the three concepts, plays a fundamental role. A notable contribution in this domain is the study by [32], which introduces the DistVAE model for sequential recommendation tasks. This model was empirically evaluated using four prominent datasets: ML-latest, ML-1M, MTweeting, and PEEK. The authors benchmarked DistVAE against eight established methods in the literature, namely, GRU4Rec, Caser, BERT4Rec, SVAE, ACVAE, FedFast, FedRec+, and FMSS. To ensure comprehensive performance assessment, the study employed three widely recognized evaluation metrics: REC, normalized discounted cumulative gain (NDCG), and mean reciprocal rank (MRR). The experimental results demonstrated that DistVAE consistently outperformed competing models across most evaluation settings. Specifically, on the ML-latest dataset, DistVAE achieved superior performance across all the metrics except Recall, where ACVAE slightly outperformed it. For the MTweeting dataset, DistVAE maintained a dominant position across all the evaluation metrics. Finally, regarding the PEEK dataset, the proposed model exhibited leading performance across all criteria except NDCG, where ACVAE marginally outperformed it. These findings underscore the robustness and adaptability of the DistVAE framework across diverse data environments. However, the occasional outperformance of ACVAE in certain recall and NDCG benchmarks suggests that while DistVAE is a highly competitive model, further refinement may be required to optimize its sensitivity to specific ranking-based metrics.

TABLE II Methods and evaluation metric results extracted from the literature.

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
[48]	CCV	DiMSC	0.222	0.199	0.255	0.127																		
		RAMSC	0.259	0.23	0.29	0.146																		
		MLFA	0.218	0.193	0.244	0.158																		
		DAMC	0.259	0.227	0.267	0.2																		
		AGARL	0.275	0.279	0.296	0.292																		
	MSRC-V1	DiMSC	0.708	0.582	0.709	0.556																		
		RAMSC	0.69	0.639	0.719	0.589																		
		MLFA	0.657	0.649	0.692	0.633																		
		DAMC	0.51	0.435	0.543	0.525																		
		AGARL	0.864	0.79	0.872	0.865																		
	Reuters	DiMSC	0.482	0.308	0.517	0.321																		
		RAMSC	0.512	0.309	0.526	0.351																		
		MLFA	0.358	0.266	0.508	0.337																		
		DAMC	0.238	0.061	0.262	0.237																		
		AGARL	0.537	0.318	0.547	0.552																		
	MNIST	DiMSC	0.32	0.187	0.326	0.205																		
		RAMSC	0.782	0.745	0.782	0.676																		
		MLFA	0.697	0.643	0.68	0.635																		
		DAMC	0.755	0.635	0.755	0.77																		
		AGARL	0.868	0.843	0.876	0.856																		
	Caltech101-20	DiMSC	0.28	0.342	0.571	0.466																		
		RAMSC	0.488	0.659	0.767	0.732																		
		MLFA	0.597	0.588	0.693	0.516																		
		DAMC	0.357	0.49	0.67	0.302																		
		AGARL	0.61	0.672	0.773	0.746																		
	VOC	DiMSC	0.488	0.496	0.517	0.498																		
		RAMSC	0.527	0.546	0.539	0.523																		
		MLFA	0.558	0.556	0.544	0.538																		
		DAMC	0.56	0.552	0.583	0.601																		
		AGARL	0.607	0.615	0.628	0.62																		

Ref.	Dataset	Metrics / Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
[51]		NN/FF-MAGO					83.89																	
		NN/FF-MINO					84.1																	
		NN/FF-OVL					85.14																	
		C4.5/FF-MAGO					39.54																	
		C4.5/FF-MINO					39.54																	
		C4.5/FF-OVL					40.37																	
		SVM/FF-MAGO					19.03																	
		SVM/FF-MINO					18.61																	
		SVM/FF-OVL					19.66																	
		LR/FF-MAGO					66.94																	
		LR/FF-MINO					66.52																	
		LR/FF-OVL					67.78																	
[49]	Image-MNIST	k-means		0.432	0.513				0.327															
		ClusterGAN		0.894	0.901				0.881															
		InfoGAN		0.816	0.839				0.852															
		GAN-SOM		0.798	0.855				0.724															
		eClusterGAN		0.815	0.926				0.8															
	Image-Fashion-MNIST	k-means		0.326	0.423				0.227															
		ClusterGAN		0.614	0.603				0.505															
		InfoGAN		0.559	0.621				0.443															
		GAN-SOM		0.569	0.663				0.416															
		eClusterGAN		0.614	0.711				0.53															
	Discrete data - MNIST	k-means		0.541	0.534				0.476															
		ClusterGAN		0.734	0.811				0.657															
		InfoGAN		0.578	0.645				0.437															
		GAN-SOM		0.787	0.846				0.669															
		eClusterGAN		0.801	0.875				0.71															
	Discrete data - Fashion-MNIST	k-means		0.503	0.512				0.489															
		ClusterGAN		0.743	0.776				0.626															
		InfoGAN		0.753	0.728				0.634															

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		GAN-SOM		0.765	0.718				0.639															
		eClusterGAN		0.789	0.82				0.79															
[54]	early-twitter	MF	0.603																					
		DW	0.579																					
		JNET	0.601																					
		STMF	0.612																					
		DisVAE	0.647																					
		DisVAEF	0.663																					
	late-twitter	MF	0.625																					
		DW	0.583																					
		JNET	0.637																					
		STMF	0.643																					
		DisVAE	0.696																					
		DisVAEF	0.718																					
	Synthetic	MF	0.639																					
		DW	0.552																					
		JNET	0.595																					
		STMF	0.655																					
		DisVAE	0.712																					
		DisVAEF	0.733																					
[52]	DS-1	DT	94.8				95.25																	
		GB	95.49				96.18																	
		KNN	94.82				95.93																	
		RF	95.35				96.25																	
		SVM	93.96				93.67																	
	DS-2	GB	94.32				92.25																	
		DT	92.1				86.77																	
		SVM	94.14				91.88																	
		RF	95.76				94.25																	
		KNN	92.21				90.61																	
	DS-3	DT	82.51				84.97																	

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		GB	83.76				87.23																	
		KNN	81.16				84.95																	
		RF	82.89				85.84																	
		SVM	82.4				87.88																	
	DS-4	KNN	93.76				93.97																	
		SVM	95.2				95.69																	
		DT	95.73				96.14																	
		RF	97.8				97.85																	
		GB	97.52				97.65																	
[53]		NB-FSS					98.148	98.148		0.028														
		NB-ECT-subspace					95.679	95.975		0.056														
		NB-ECT-noise					96.914	97.05		0.019														
		NB-ECT-combined					98.148	98.452		0.019														
		NB-LPP					92.593	93.75		0.074														
		NB-NPE					91.358	92.5		0.093														
		NB-IsoP					93.827	95		0.056														
		C4.5-FSS					95.062	94.19		0.102														
		C4.5-ECT-subspace					94.444	94.737		0.074														
		C4.5-ECT-noise					95.679	94.801		0.093														
		C4.5-ECT-combined					96.914	96.615		0.056														
		C4.5-LPP					90.123	90.966		0.12														
		C4.5-NPE					88.889	93.705		0.102														
		C4.5-IsoP					91.975	92.315		0.139														
		SVM-FSS					81.481	88.889		0.028														
		SVM-ECT-subspace					84.568	89.251		0.074														
		SVM-ECT-noise					85.802	89.389		0.093														
		SVM-ECT-combined					88.272	92.557		0.037														
		SVM-LPP					82.099	86.645		0.111														
		SVM-NPE					79.63	87.162		0.046														
		SVM-IsoP					80.864	86.903		0.065														

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		LR-FSS					69.136	75.676		0.204														
		LR-ECT-subspace					74.074	79.734		0.176														
		LR-ECT-noise					75.926	81.457		0.157														
		LR-ECT-combined					76.543	81.579		0.167														
		LR-LPP					66.049	72.546		0.241														
		LR-NPE					67.284	73.649		0.231														
		LR-IsoP					64.815	73.091		0.185														
[58]	Cornell	node2vec	33.85	6.65		72.63					70.99													
		GraRep	31.79	8.8		47.42					43.87													
		SNE	41.08	11.11		51.51					52.99													
		VGAE	36.72	7.77		85.99					82.94													
		ARVGA	38.21	10.26		85.54					83.92													
		ArmGAN$_m$	54.36	21.07		91.9					88.56													
		ArmGAN$_d$	48.2	15.24		92.55					91.32													
	Texas	node2vec	47.54	4.49		57.31					55.3													
		GraRep	36.72	12.43		47.35					44.42													
		SNE	41.53	12.63		50.98					51.57													
		VGAE	48.35	8.52		85.71					80.88													
		ARVGA	41.48	7.28		81.08					76.45													
		ArmGAN$_m$	60.66	18.42		92.29					89.16													
		ArmGAN$_d$	56.28	13.55		93.23					89.7													
	Washington	node2vec	37.33	2.94		60.89					56.63													
		GraRep	31.36	5.18		47.89					45.57													
		SNE	48.8	17.43		49.51					49.89													
		VGAE	43.73	9.03		80.55					75.54													
		ARVGA	43.66	12.6		83.66					77													
		ArmGAN$_m$	60.83	25.91		86.25					80.66													
		ArmGAN$_d$	60.82	25.82		85.98					81.47													
	Wisconsin	node2vec	49.62	7.86		70.75					69.43													
		GraRep	33.24	8.02		47.73					45.43													

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		SNE	55.3	19.84		52.23					54.07													
		VGAE	43.28	9.31		85.68					83.3													
		ARVGA	42.81	11.22		76.25					68.78													
		ArmGAN$_m$	56.49	19.72		92.05					89.64													
		ArmGAN$_d$	58.01	19.94		92.97					91.01													
	Cora	node2vec	56.3	42.02		74.61					77.39													
		GraRep	48.29	35.46		52.89					55.31													
		SNE	39.44	16.28		76.85					84.68													
		VGAE	57.06	42.92		93.51					92.38													
		ARVGA	64.08	44.95		92.99					92.8													
		ArmGAN$_m$	76.11	58.43		94.52					94.29													
		ArmGAN$_d$	74.04	58.22		95.26					94.99													
	Citeseer	node2vec	40.76	12.99		68.09					67.31													
		GraRep	31.2	9.61		64.11					69.03													
		SNE	31.17	7.31		75.3					83.09													
		VGAE	53.46	27.93		92.66					91.44													
		ARVGA	43.5	22.72		93.48					92.41													
		ArmGAN$_m$	70.18	44.56		96.13					95.46													
		ArmGAN$_d$	67.77	42.89		96.79					96.81													
	Pubmed	node2vec	65.56	25.02		76.97					78.03													
		GraRep	54.43	17.76		48.26					46.33													
		SNE	65.13	25.61		78.73					75.52													
		VGAE	58.64	17.83		94.86					94.46													
		ARVGA	58.76	18.4		96.29					96.11													
		ArmGAN$_m$	71.55	33.7		95.64					95.64													
		ArmGAN$_d$	70.96	32.91		96.34					96.7													
[56]	Dataset1/ResNet-50-IO	k-means	0.9022									379837	145.38	0.6847	0.62									
		ACS	0.9502									384091	154.61	0.7437	0.69									
	Dataset1/ResNet-50-AS	k-means	0.9101									335414	157.86	0.7183	0.66									
		ACS	0.9597									339441	168.38	0.769	0.7									

Ref.	Dataset	Metrics / Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
	Dataset2/ResNet-50-IO	k-means	0.2446									148483	236.42	0.244	0.19									
		ACS	0.2643									150272	246.56	0.264	0.21									
	Dataset2/ResNet-50-AS	k-means	0.2816									13272	253.42	0.2816	0.24									
		ACS	0.3025									134145	260.11	0.302	0.26									
	Dataset3/ResNet-50-IO	k-means	N/A									57279	111.92	0.891	N/A									
		ACS	N/A									58392	124.83	0.915	N/A									
	Dataset3/ResNet-50-AS	k-means	N/A									38577	93.13	0.874	N/A									
		ACS	N/A									39826	106.95	0.904	N/A									
[46]	Iris	K-means	C= 89.3, P=47.3	C= 0.775, P=0.363												N/A	N/A							
		GMM	C= 96.7, P=52.9	C= 0.898, P=0.395												N/A	N/A							
	MNIST	K-means	C= 52.8, P=22.5	C= 0.501, P=0.291												51.10%	0.00							
		GMM	C= 53.7, P=24	C= 0.525, P=0.303												53.20%	0.10							
	TIMIT	GMM-UBM	C= 63.6, P=1.2	C= 0.598, P=0.171												60.40%	0.00							
[27]	Cora	GAE	0.679	0.504		0.726		0.673	0.443															
		VGAE	0.696	0.519		0.71		0.679	0.473															
		ARGA	0.708	0.517		0.72		0.694	0.472															
		ARVGA	0.68	0.52		0.687		0.658	0.462															
		EVGAE	0.678	0.502		0.685		0.666	0.45															
		GNAE	0.712	0.543		0.737		0.692	0.504															
		VGNAE	0.717	0.544		0.718		0.695	0.508															
		ARVGNA	0.734	0.571		0.744		0.718	0.54															
	CiteSeer	GAE	0.455	0.258		0.549		0.423	0.134															
		VGAE	0.608	0.373		0.587		0.555	0.338															
		ARGA	0.447	0.255		0.558		0.419	0.118															
		ARVGA	0.595	0.368		0.587		0.549	0.326															
		EVGAE	0.551	0.305		0.565		0.523	0.261															
		GNAE	0.599	0.382		0.593		0.556	0.345															

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		VGNAE	0.562	0.335		0.603		0.542	0.265															
		ARVGNA	0.623	0.388		0.613		0.585	0.36															
	PubMed	GAE	0.653	0.249		0.674		0.64	0.247															
		VGAE	0.661	0.258		0.674		0.647	0.263															
		ARGA	0.662	0.265		0.681		0.649	0.264															
		ARVGA	0.662	0.255		0.674		0.647	0.264															
		EVGAE	0.659	0.259		0.673		0.646	0.261															
		GNAE	0.674	0.278		0.687		0.66	0.283															
		VGNAE	0.674	0.275		0.684		0.661	0.284															
		ARVGNA	0.677	0.283		0.69		0.664	0.29															
[23]	YelpCHI	Average																0.587						
		Popular																0.54						
		Random																0.584						
		Trial																0.694						
		PoisonT																0.661						
		MetaC																0.864						
	Movies	Average																0.392						
		Popular																0.384						
		Random																0.398						
		Trial																0.409						
		PoisonT																0.401						
		MetaC																0.828						
[59]	ML-1M	RandomAttack																0.06	0.0055					
		Bandwagon																0.0595	0.0047					
		AUSH																0.0764	0.0049					
		FedRecAttack																0.1049	0.0221					
		RAPU-G																0.0922	0.0287					
		GTA																0.0824	0.0199					
		PoisonRec																0.0842	0.0274					
		InfoAtk																0.0958	0.0167					

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
	Douban	RandomAttack																0.0003	-0.0136					
		Bandwagon																0.0002	-0.0249					
		AUSH																0.0029	-0.0126					
		FedRecAttack																0.0044	-0.0145					
		RAPU-G																N/A	N/A					
		GTA																0.0056	-0.0109					
		PoisonRec																0.006	-0.0113					
		InfoAtk																0.0113	-0.0088					
	Epinions	RandomAttack																0.0182	-0.0154					
		Bandwagon																0.0298	-0.0201					
		AUSH																0.0236	-0.0136					
		FedRecAttack																0.0627	-0.0112					
		RAPU-G																–	–					
		GTA																0.0232	-0.0154					
		PoisonRec																0.0292	-0.015					
		InfoAtk																0.0775	-0.0024					
[68]	Edinburgh	C-ILP						0.509																
		TRED						0.577																
		DeepTrip						0.633																
		GC-TripRec						0.671																
	Glasgow	C-ILP						0.634																
		TRED						0.633																
		DeepTrip						0.673																
		GC-TripRec						0.735																
	Osaka	C-ILP						0.463																
		TRED						0.602																
		DeepTrip						0.617																
		GC-TripRec						0.682																
	Toronto	C-ILP						0.58																
		TRED						0.668																

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		DeepTrip						0.69																
		GC-TripRec						0.748																
[65]	MovieLens 1M	FNCF																0.6576	0.3802					
		FNCF-Single																0.6575	0.3821					
		FNCF-Multi																0.6571	0.3824					
	MovieLens 100K	FNCF																0.7158	0.4216					
		FNCF-Single																0.7137	0.4216					
		FNCF-Multi																0.7094	0.4188					
[25]	ML10M-FX	Without AT																0.0228	0.0195					
		RL-GEN																0.0324	0.0222					
		RandomAT																0.023	0.0195					
		TA-AT40																0.0583	0.0195					
		TA-AT70																0.0854	0.0341					
		TA-AT100																0.052	0.0209					
		PolicyNetwork																0.0665	0.0258					
		CopyAT-Masking																0.0227	0.0195					
		CopyAT-Length																0.0434	0.0177					
		CopyAttack																0.1103	0.0425					
		CopyAT+(Two)																0.1123	0.0402					
		CopyAT+(Joint)																0.1313	0.0516					
	ML20M-NF	Without AT																0.0043	0.0013					
		RL-GEN																0.039	0.0226					
		RandomAT																0.005	0.0015					
		TA-AT40																0.0405	0.0133					
		TA-AT70																0.0402	0.0132					
		TA-AT100																0.0006	0.0002					
		PolicyNetwork																N/A	N/A					
		CopyAT-Masking																0.0045	0.0001					
		CopyAT-Length																0.0018	0.0005					
		CopyAT																0.124	0.0609					

Ref.	Dataset	Metrics / Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		CopyAT+(Two)																0.1273	0.0627					
		CopyAT+(Joint)																0.1332	0.0656					
[3]	MovieLens-1M	DAAKG-Rdrop																		WRMF-SGD=0.929 ItemAE=0.939 GOAT=0.959 Average=0.974 BandWagon=0.974 mixed attack=0.964				
		DAAKG-S																		WRMF-SGD=0.919 ItemAE=0.924 GOAT=0.939 Average=0.964 BandWagon=0.964 Mixed attack=0.959				
	Book-Crossing	DAAKG-Rdrop																		WRMF-SGD=0.938 ItemAE=0.948 GOAT=0.974 Average=0.980 BandWagon=0.990 mixed attack=0.984				

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		DAAKG-S																		WRMF-SGD=0.906 ItemAE=0.918 GOAT=0.959 Average=0.9796BandWagon=0.979 mixed attack=0.964				
	Nowplaying	DAAKG-Rdrop																		WRMF-SGD=0.952 ItemAE=0.947 GOAT=0.979 Average=0.989 BandWagon=0.989 mixed attack=0.989				
		DAAKG-S																		WRMF-SGD=0.927 ItemAE=0.936 GOAT=0.959 Average=0.974 BandWagon=0.979 mixed attack=0.9799				
[70]	Ciao	APR				0.0482	0.1432												0.1104					
		BUIR				0.0536	0.1653												0.1239					
		IRGAN				0.0437	0.1252												0.1084					
		CFGAN				0.0547	0.1599												0.1238					
		FairGAN				0.0556	0.1621												0.1254					
		StuGAN				0.0574	0.1692												0.1303					
	LastFM	APR				0.1431	0.1936												0.2132					

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		BUIR				0.1368	0.1978												0.2249					
		IRGAN				0.1039	0.1624												0.1713					
		CFGAN				0.1416	0.1925												0.2146					
		FairGAN				0.1446	0.1969												0.2235					
		StuGAN				0.1502	0.2032												0.2308					
[69]		CF					0.2												0.11					
		HCF					0.2												0.13					
		DNN-BWMRB					0.4												0.28					
[5]	MovieLens	BPR				0.1776	0.0742												0.4416					
		CDL				0.1412	0.0726												0.4375					
		CVAE				0.1532	0.0874												0.4465					
		CVAE-B				0.1825	0.0915												0.4502					
		IRGAN				0.1459	0.0753												0.4391					
		VAE-AR				0.1964	0.0979												0.4706					
		CLVAE				0.1981	0.0969												0.4701					
		CFGAN				0.1785	0.0927												0.4562					
		MVAE				0.1897	0.0942												0.4643					
		CAF				0.2139	0.1069												0.4856					
	CiteULike	BPR				0.0885	0.0824												0.3201					
		CDL				0.1008	0.1047												0.3514					
		CVAE				0.1461	0.1291												0.375					
		CVAE-B				0.1627	0.1482												0.4021					
		IRGAN				0.1015	0.114												0.3502					
		VAE-AR				0.1322	0.1172												0.3628					
		CLVAE				0.1447	0.1256												0.3724					
		CFGAN				0.1412	0.1026												0.3583					
		MVAE				0.1546	0.1056												0.362					
		CAF				0.1795	0.1649												0.4587					
	LastFM	BPR				0.2853	0.2614												0.5249					
		CDL				0.3102	0.294												0.5612					

Ref.	Dataset	Metrics / Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		CVAE				0.3361	0.3148												0.6079					
		CVAE-B				0.3543	0.3274												0.6312					
		IRGAN				0.3068	0.3002												0.5776					
		VAE-AR				0.3615	0.3366												0.6328					
		CLVAE				0.3671	0.3006												0.6099					
		CFGAN				0.3226	0.3117												0.6142					
		MVAE				0.3418	0.302												0.6035					
		CAF				0.3804	0.3578												0.6618					
[37]	Chinese	Feedrec				0.54	0.68	0.3																
		GACF				0.58	0.69	0.32																
		KEHB				0.48	0.67	0.28																
		MACR				0.56	0.68	0.31																
		Proposed				0.61	0.72	0.33																
	English	Feedrec				0.62	0.76	0.29																
		GACF				0.72	0.82	0.33																
		KEHB				0.62	0.75	0.29																
		MACR				0.69	0.8	0.32																
		Proposed				0.74	0.84	0.36																
	Adressa	Feedrec				0.51	0.65	0.23																
		GACF				0.69	0.81	0.32																
		KEHB				0.51	0.64	0.24																
		MACR				0.58	0.79	0.31																
		Proposed				0.75	0.82	0.35																
	Digg	KEHB				0.59	0.78	0.39																
		Feedrec				0.65	0.79	0.41																
		MACR				0.67	0.79	0.42																
		GACF				0.69	0.8	0.43																
		Proposed				0.72	0.83	0.44																
[38]	Yelp (Pittsburgh)	N2VSCDNNR				0.0153	0.0818											0.201						
		Metapath2vec++				0.0095	0.0326											0.1164						
		BiNE				0.0115	0.0399											0.1352						

Ref.	Dataset	Metrics / Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		CoFactor				0.0144	0.0735											0.1832						
	Yelp (Madison)	N2VSCDNNR				0.0166	0.1042											0.217						
		Metapath2vec++				0.0101	0.0341											0.1214						
		BiNE				0.0128	0.0542											0.1512						
		CoFactor				0.016	0.0951											0.2088						
	Amazon	N2VSCDNNR				0.0093	0.0608											0.124						
		Metapath2vec++				0.0038	0.0257											0.0643						
		BiNE				0.0059	0.0389											0.0875						
		CoFactor				0.0087	0.0575											0.1141						
	MovieLens	N2VSCDNNR				0.2933	0.2865											0.935						
		Metapath2vec++				0.281	0.275											0.8957						
		BiNE				0.1721	0.1236											0.7108						
		CoFactor				0.281	0.275											0.8957						
[1]	Movielens 100K	NMF																			0.7519	0.9499		
		PMF																			0.7286	0.9226		
		BPMF																			0.6976	0.888		
		SVD++																			0.7109	0.9052		
		ReDa																			0.7153	0.9114		
		HRSA																			0.7051	0.8961		
		HCRDa																			0.609	0.7879		
	MovieTweetings 10K	NMF																			1.2373	1.7247		
		PMF																			1.341	1.7526		
		BPMF																			1.2898	1.7105		
		SVD++																			1.1397	1.5296		
		ReDa																			-	-		
		HRSA																			2.2989	2.4629		
		HCRDa																			0.617	0.8588		
	FilmTrust	NMF																			0.6476	0.8539		
		PMF																			0.6948	0.9226		
		TrustMF																			0.6342	0.828		

Ref.	Dataset	Metrics / Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		SVD++																			0.6242	0.8049		
		TrustSVD																			0.6188	0.7943		
		HRSA																			0.6256	0.8103		
		HCRDa																			0.552	0.7225		
[4]		CF(DSP)																				0.7591		
		MR(DSP)																				0.7382		
		CF(NN)																				0.7274		
		MR(NN)																				0.7152		
[39]	LDOS-CoMoDa	CBRS (SOM)																			0.999	0.7653	2.087	
	InCarMusic	CBRS (SOM)																			0.863	0.9873	2.083	
	Apps in Frappe	CBRS (SOM)																			0.378	0.1862	1.077	
	POI in STS	CBRS (SOM)																			0.709	0.8092	0.977	
	Hotels in Trip Advisor	CBRS (SLINK)																			1.252	1.5847	1.512	
	Drug Reviews	CBRS (SOM)																			0.654	0.7334	1.114	
	Apple Store	CBRS (SOM)																			1.335	2.1928	1.376	
[43]		CF KNN + baseline																				0.8535		
		Co-clustering (K-means)																				0.9131		
		NMF																				0.8746		
		slope-one																				0.8864		
[19]		TSTDAE				0.0772												0.173	0.1419					
		NCF				0.0551												0.0295	0.0991					
		LRML				0.0363												0.2031	0.0845					
		JRL				0.0428												0.0911	0.1069					
		CML				0.0674												0.0571	0.0614					
		UserKNN				0.0659												0.1222	0.1208					
		ItemKNN				0.0685												0.0899	0.102					
		DDTCDR				0.0666												0.1101	0.1021					

Ref.	**Dataset**	**Metrics** **Methods**	**Acc**	**NMI**	**Pur**	**PRE**	**REC**	**F1**	**ARI**	**FPR**	**AUC**	**WCSS**	**CHI**	**A_usu**	**α**	**TC**	**Test AE**	**HR@**	**NDCG@**	**Attack type**	**MAE**	**RMSE**	**MSE**	**MRR**
[40]		LSTM				0.282	0.75																	
		CNN				0.25	0.65																	
		Auto-Encoder				0.27	0.66																	
		GMDH				0.35	0.81																	
[41]		LSTM	0.88			0.86	0.83																	
		Attention	0.9			0.88	0.85																	
		Proposed	0.96			0.94	0.94																	
[42]	Movielens 100 k	GI-AAE				0.48	0.09	0.151																
		AVAE				0.465	0.086	0.145																
		CDAE				0.46	0.082	0.139																
		SLIM				0.36	0.128	0.188																
		LFM				0.346	0.123	0.18																
		UserCF				0.26	0.1	0.144																
		ItemCF				0.29	0.11	0.159																
	Movielens 1M	GI-AAE				0.41	0.102	0.163																
		AVAE				0.38	0.082	0.134																
		CDAE				0.39	0.08	0.132																
		SLIM				0.35	0.09	0.143																
		LFM				0.336	0.08	0.129																
		UserCF				0.27	0.07	0.111																
		ItemCF				0.3	0.06	0.1																
	FilmTrust	GI-AAE				0.44	0.55	0.488																
		AVAE				0.43	0.46	0.444																
		CDAE				0.42	0.425	0.422																
		SLIM				0.4	0.52	0.452																
		LFM				0.39	0.51	0.442																
		UserCF				0.38	0.48	0.424																
		ItemCF				0.35	0.44	0.389																
[44]	Last.Fm	CKE					0.344												0.173					

Ref.	Dataset	Metrics / Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		KGAT					0.339												0.160					
		KGIN					0.361												0.176					
		KGCL					0.38												0.194					
		CSEKG					0.395												0.197					
	Amazon-book	CKE					0.173												0.084					
		KGAT					0.163												0.077					
		KGIN					0.195												0.093					
		KGCL					0.187												0.091					
		CSEKG					0.204												0.098					
	MovieLens-1M	CKE					0.296												0.207					
		KGAT					0.32												0.221					
		KGIN					0.322												0.239					
		KGCL					0.338												0.233					
		CSEKG					0.357												0.252					
[71]	Synthetic	OCCUD									0.855													
		GCUD									0.502													
		NCUD									0.464													
	Movielens	OCCUD									0.85													
		GCUD									0.492													
		NCUD									0.449													
	Amazon	OCCUD									0.84													
		GCUD									0.518													
		NCUD									0.469													
	Yelp	OCCUD									0.628													
		GCUD									0.51													
		NCUD									0.509													
[32]	ML-latest	GRU4Rec					0.0804												0.0515					0.1207
		Caser					0.1068												0.0751					0.1288
		BERT4Rec					0.0877												0.0609					0.1282

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		SVAE					0.1203												0.0889					0.1484
		ACVAE					0.1449												0.1066					0.1793
		FedFast					0.0055												0.0031					0.0042
		FedRec++					0.0009												0.0008					0.0017
		FMSS					0.081												0.0511					0.1224
		DistVAE					0.1404												0.1103					0.1964
	ML-1M	GRU4Rec					0.1309												0.1181					0.221
		Caser					0.1857												0.1639					0.2812
		BERT4Rec					0.1543												0.1293					0.2302
		SVAE					0.2073												0.176					0.2994
		ACVAE					0.2438												0.2122					0.3484
		FedFast					0.0026												0.0054					0.007
		FedRec++					0.0005												0.0003					0.0002
		FMSS					0.1314												0.1184					0.2207
		DistVAE					0.2454												0.2171					0.3619
	MTweeting	GRU4Rec					0.155												0.0751					0.0732
		Caser					0.2088												0.1347					0.1326
		BERT4Rec					0.2744												0.1442					0.151
		SVAE					0.2775												0.1591					0.1603
		ACVAE					0.278												0.1608					0.1617
		FedFast					0.0014												0.0026					0.0037
		FedRec++					0.0008												0.0003					0.0002
		FMSS					0.1552												0.0758					0.0739
		DistVAE					0.2801												0.1652					0.1703
	PEEK	GRU4Rec					0.0951												0.0671					0.0583
		Caser					0.1208												0.081					0.0719
		BERT4Rec					0.1761												0.1074					0.0872
		SVAE					0.1601												0.1109					0.0872
		ACVAE					0.1876												0.1213					0.1068
		FedFast					0.0041												0.0035					0.0035

Ref.	Dataset	Metrics / Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		FedRec++					0.0015												0.0003				0.0003	
		FMSS					0.0959												0.0677				0.058	
		DistVAE					0.193												0.12				0.1071	
[31]	Beauty	Baseline/Pure																0.0575	0.0198					0.0348
		Baseline/Random																0.0707	0.024					0.0367
		Baseline/Popular																0.0691	0.024					0.0354
		Baseline/RecommPoison																0.0787	0.0266					0.0382
		Baseline/SeqPoison																0.1059	0.0381					0.0446
		ClusterPoison/K-means(SeqPoison)																0.1269	0.0477					0.0489
		ClusterPoison/Clique(SeqPoison)																0.0806	0.0284					0.0346
		ClusterPoison/DBSCAN(SeqPoison)																0.141	0.0494					0.0522
		ClusterPoison/K-means(RecommPoison)																0.1221	0.0471					0.0477
		ClusterPoison/Clique(RecommPoison)																0.1187	0.0436					0.046
		ClusterPoison/DBSCAN(RecommPoison)																0.1209	0.0449					0.0472
	Sports	Baseline/Pure																0.0533	0.019					0.0328
		Baseline/Random																0.065	0.0241					0.0361
		Baseline/Popular																0.0474	0.0178					0.0324
		Baseline/RecommPoison																0.0672	0.0251					0.0359
		Baseline/SeqPoison																0.0592	0.0212					0.0361
		ClusterPoison/K-means(SeqPoison)																0.0874	0.0345					0.0438
		ClusterPoison/Clique(SeqPoison)																0.0954	0.0369					0.0437
		ClusterPoison/DBSCAN(SeqPoison)																0.0886	0.0342					0.0409

Ref.	Dataset	Metrics / Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		ClusterPoison/K-means(RecommPoison)																0.1252	0.05					0.0516
		ClusterPoison/Clique(RecommPoison)																0.1219	0.0506					0.0519
		ClusterPoison/DBSCAN(RecommPoison)																0.0785	0.0318					0.041
[72]	MovieLens-100k	Supervised/Pop-SAD						Random=96.9 Average=93.7 Bandwagon=93.3 AvgTargetShift=94.8 AvgNoiseInject=94.9 AoP=99.4 PIA=100																
		Supervised/SpDetector						Random=99.9 Average=99.7 Bandwagon=99.9 AvgTargetShift=99.4 AvgNoiseInject=99.7 AoP=99.8 PIA=100																
		Unsupervised/PCA-VarSelect						Random=97 Average=96 Bandwagon=95.5 AvgTargetShift=96.6 AvgNoiseInject=96.2 AoP=19.8 PIA=0.0																

Ref.	Dataset	Metrics / Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		Unsupervised/ ProMethod						Random=100 Average=100 Bandwagon=99.9 AvgTargetShift=100 AvgNoiseInject=100 AoP=99.3 PIA=99.3																
	Netflix	Supervised/Pop -SAD						Random=98.5 Average=99 Bandwagon=98.5 AvgTargetShift=98.5 AvgNoiseInject=99 AoP=88.9 PIA=100																
		Supervised/Sp Detector						Random=99.9 Average=99.9 Bandwagon=99.9 AvgTargetShift=99.9 AvgNoiseInject=99.9 AoP=99.6 PIA=99.9																
		Unsupervised/P CA-VarSelect						Random=98.5 Average=97.2 Bandwagon=97 AvgTargetShift=97.4 AvgNoiseInject=97 AoP=68.3 PIA=0																

Ref.	Dataset	Metrics Methods	Acc	NMI	Pur	PRE	REC	F1	ARI	FPR	AUC	WCSS	CHI	A_usu	α	TC	Test AE	HR@	NDCG@	Attack type	MAE	RMSE	MSE	MRR
		Unsupervised/ProMethod						Random=100 Average=100 Bandwagon=100 AvgTargetShift=100 AvgNoiseInject=100 AoP=99.5 PIA=99.5																

ARI=Adjusted Rand index, FPR=False Positive Rate, AUC=Area Under the Curve, α=Krippendorff's Alpha-Reliability, C=clean, P= Poison, AE= Adversarial Examples, DSP=Data Sparsity, NN=nearest neighbor, MR =Mixed recommendation, A_usup=Accuracy unsupervised, CHI =Calinski–Harabasz Index, HR@=Hit ratio, MAE= Mean absolute error, RMSE= Root Mean Squared Error, MSE= Mean Squared Error, TC=Test Clean, NN=Naive Bayes, LR= Logistic Regression, MAGO=Majority, MINO=Minority, OVL=Overall, RL–GEN=RL–Generative, attack=AT, TA=Target

6.7 Overview of Integrating Clustering, RSs, and Adversarial Learning

Table III explains that the document critically examines four studies addressing challenges in RSs. Study [32] proposes DistVAE, a distributed variational autoencoder, to mitigate data sparsity and gradient instability in sequential recommendations, although it lacks clarity in handling long-term dependencies and scalability. Study [72] introduces a three-phase framework for shilling attack detection, enhancing adaptability but overlooking real-world deployment challenges and evolving attack sophistication. Study [71], the OCCUD method leverages online clustering for corrupted user detection, offering real-time adaptability but ambiguously defining "corruption" and neglecting scalability for large platforms. Study [31] ClusterPoison demonstrated how clustering-enhanced poisoning attacks exploit RSs vulnerabilities with minimal fake users but ignore ethical implications and defensive countermeasures. Collectively, these studies highlight innovative architectures (e.g., DistVAE, OCCUD) and refined attack strategies but suffer from methodological gaps such as undefined evaluation metrics, scalability limitations, and insufficient ethical considerations. While advancing theoretical solutions, broader empirical validation, interdisciplinary integration (e.g., ethics, adversarial defense), and industrial-scale testing are essential to translate these advancements into robust, real-world RSs frameworks.

TABLE III Overview of integrated clustering, RSs, and adversarial learning

Ref.	Authors	Challenges	Methods	solved Challenges
[32]	L. Li *et al.*	- The paper talks about some of the problems that come up when trying to use a distributed version of Variational Autoencoder for sequential recommendation. - First, there is not much data on a single device in a distributed setup, which makes the gradients used to update the global model more random. - Second, people who utilize recommendation systems have very different preferences, which might make things more random and cause global model updates to be less stable. - Finally, many generative approaches still use GRU, which has trouble modelling long-term dependencies in user sequences.	DistVAE	✓
[72]	*Fei Zhang et al.*	- The article addresses a problem with RSs; they can be manipulated by introducing hand-crafted profiles to mislead them, even when only a small number of attacked profiles are present. - It notes that within the recent methods of profile classification, many methods are performance oriented for certain attack scenarios, limiting the generalizability of these methods.	The suggested shilling attack detection approach comprises three steps: classifying the type of attack, finding the suspicious profile, and finding the assault profile.	✓
[71]	X. Dai *et al.*	This paper investigated important challenges in the domain of corrupted user detection for Collaborative RSs. Corrupted behaviors have the potential to corrupt estimates of user preferences, which leads to invalid inferences regarding user relationships and poor recommendations.	OCCUD	✓
[31]	Y. Wang *et al.*	- The article addresses the potential threat of manipulating many fake users (i.e., many millions of users) on recommendation platforms in large settings where it is physically impossible for attackers to go undetected. - The study underlines the need for better defenses against poisoning attacks, when the fake users are, theoretically, extremely limited as any one fake user can easily cause harm to the recommendation.	The paper introduces a framework clustering-based scheme for generating fake users, which can be integrated into various poisoning attacks against deep learning-based RSs.	✓

7. CONCLUSION

In this work, we present a comprehensive study on the intersection between RSs, clustering concepts, and adversarial learning. The topics were discussed from several areas on the basis of the motivations of the extracted research, the challenges faced by researchers, and the recommendations and future work that the researchers wanted to work on in the future. The results of this work revealed that 51 research studies within four reliable databases and taxonomy have four main categories: 1) those based on RSs and clustering; 2) those based on clustering and adversarial learning; 3) those based on RSs and adversarial learning; and 4) Integration based on RSs, clustering, and adversarial learning. In addition, an analysis of studies was explained based on cooperation between countries, the most frequently repeated keywords in the research, and other analysis methods specific to the research. Additionally, the intersection that occurs between the RSs and clustering concepts produces more accurate results. Since the clustering approach is based on data clustering, determining the correct number of clusters is essential. Therefore, recommendation systems are built based on the data provided, which must be accurate. The more correct the clusters are, the more accurate the recommendation and vice versa. As a result, we found the following:

First, after counting the data used in each study, we found that the research focused on data related to RSs. Most of the research on the data used, or case studies, has focused on entertainment and commerce. Other areas, such as healthcare and finance, were neglected. This is considered a research gap that has not been addressed, summarized, or discussed, and because challenges and limitations exist in any field, this aspect has not been considered.

Second, the other point is that the process of building any model requires building a model that is relatively robust, flexible to changes, and generalizable. Relatively, the majority and large percentage of extracted research discussed and provided the highest accuracy in the field of research, forgetting an important thing: all research that is built must be flexible and robust. In the event that the model is exposed to an attack, it must be highly defensive in its results so that it does not suffer any defects or reveal any bias in the event that it is exposed to attacks. Here, the issue of integrating adversarial learning, RSs, and clustering concepts in building any model emerged, especially after the great development of internet networks. The issue of protecting information is very important, whether from tampering with it or influencing the results and changing the recommendations. Most of the topics covered in the research were related to topics of interest to users, as this manipulation could affect the quality or accuracy of the results. This could lead to the model being biased toward incorrect results, which are reflected back to the user. Ultimately, the recommendations provided by the model are designed to meet the user's expectations and do not lead to results that may dissatisfy users or even lead to poor outcomes. The field of adversarial learning is crucial for adding RSs and clustering elements to prevent any bias or misleading results from affecting the recommendations provided. In summary, this approach opens a new research area to develop models based on adversarial learning to be robust and provide accurate results. This is because when building a model, it is exposed to attacks, regardless of the nature of the attacks, but avoids any fluctuations in the results. This opens the door to discussion and future research. It is considered a guiding principle that adversarial learning must be considered in the model-building process to ensure that the model is robust and adaptable.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Funding

There was no funding for this research study.

Acknowledgement

We would like to acknowledge the outstanding efforts of the editor and reviewers in reviewing and evaluating our research.

References

[1] B. Dong, Y. Zhu, L. Li, and X. Wu, "Hybrid Collaborative Recommendation via Dual-Autoencoder," *IEEE Access*, vol. 8, pp. 46030–46040, 2020, doi: 10.1109/ACCESS.2020.2979255.

[2] Q. Zhou and C. Huang, "A recommendation attack detection approach integrating CNN with Bagging," *Comput. Secur.*, vol. 146, 2024, doi: 10.1016/j.cose.2024.104030.

[3] Y. Hao, H. Wang, Q. Zhao, L. Feng, and J. Wang, "Detecting the adversarially-learned injection attacks via knowledge graphs," *Inf. Syst.*, vol. 125, p. 102419, 2024, doi: 10.1016/j.is.2024.102419.

[4] L. Li, Z. Zhang, and S. Zhang, "Hybrid Algorithm Based on Content and Collaborative Filtering in Recommendation System Optimization and Simulation," *Sci. Program.*, vol. 2021, 2021, doi: 10.1155/2021/7427409.

[5] F. Zhou, Y. Mo, G. Trajcevski, K. Zhang, J. Wu, and T. Zhong, "Recommendation via Collaborative Autoregressive Flows," *Neural Networks*, vol. 126, pp. 52–64, 2020, doi: 10.1016/j.neunet.2020.03.010.

[6] M. H. Hadid *et al.*, "Semantic Image Retrieval Analysis Based on Deep Learning and Singular Value Decomposition," *Appl. Data Sci. Anal.*, vol. 2024, pp. 17–31, 2024, doi: 10.58496/adsa/2024/003.

[7] M. A. Ahmed *et al.*, "Intelligent Decision-Making Framework for Evaluating and Benchmarking Hybridized Multi-Deep Transfer Learning Models: Managing COVID-19 and Beyond," *Int. J. Inf. Technol. Decis. Mak.*, Apr. 2023, doi: 10.1142/S0219622023500463.

[8] S. S. Joudar, R. A. Hamid, I. A. Zahid, G. Kou, and I. M. Sharaf, "Explainable artificial intelligence multimodal of autism triage levels using fuzzy approach-based multi-criteria decision-making and LIME," *Int. J. Fuzzy Syst.*, vol. 26, no. 1, pp. 274–303, 2024.

[9] A. S. Albahri *et al.*, "A Trustworthy and Explainable Framework for Benchmarking Hybrid Deep Learning Models Based on Chest X-Ray Analysis in CAD Systems," *Int. J. Inf. Technol. Decis. Mak.*, pp. 1–54, Jan. 2024, doi: 10.1142/S0219622024500019.

[10] L. A. E. Al-Saeedi *et al.*, "Artificial Intelligence and Cybersecurity in Face Sale Contracts: Legal Issues and Frameworks," 2024. doi: 10.58496/MJCS/2024/0012.

[11] T. Al-Quraishi, C. Keong NG, O. A. Mahdi, A. Gyasi, and N. Al-Quraishi, “Advanced Ensemble Classifier Techniques for Predicting Tumor Viability in Osteosarcoma Histological Slide Images,” *Appl. Data Sci. Anal.*, vol. 2024, pp. 52–68, 2024, doi: 10.58496/adsa/2024/006.

[12] R. Nai, R. Meo, G. Morina, and P. Pasteris, “Public tenders, complaints, machine learning and recommender systems: a case study in public administration,” *Comput. Law Secur. Rev.*, vol. 51, p. 105887, 2023, doi: https://doi.org/10.1016/j.clsr.2023.105887.

[13] F. K. H. Mihna *et al.*, “Bridging Law and Machine Learning: A Cybersecure Model for Classifying Digital Real Estate Contracts in the Metaverse,” 2025. doi: 10.58496/MJBD/2025/003.

[14] M. E. Alqaysi, A. S. Albahri, and R. A. Hamid, “Evaluation and benchmarking of hybrid machine learning models for autism spectrum disorder diagnosis using a 2-tuple linguistic neutrosophic fuzzy sets-based decision-making model,” *Neural Comput. Appl.*, vol. 36, no. 29, pp. 18161–18200, 2024.

[15] M. E. Alqaysi, A. S. Albahri, and R. A. Hamid, “Hybrid Diagnosis Models for Autism Patients Based on Medical and Sociodemographic Features Using Machine Learning and Multicriteria Decision-Making (MCDM) Techniques : An Evaluation and Benchmarking Framework,” vol. 2022, no. ii, 2022.

[16] A. A. A. Lateef, A. S. Abdalkafor, and A. A. Nafea, “Optimized KNN Algorithm for Diabetic Retinopathy Classification with PCA-Based Data Fusion and Cuckoo Search Optimization,” *J. Intell. Syst. Internet Things*, vol. 15, no. 1, pp. 122–132, 2025, doi: 10.54216/JISIoT.150110.

[17] A. S. Abdalkafor and K. M. A. Alheeti, “K-Nearest Neighbor Algorithm for Efficient Heart Disease Classification System,” in *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*, 2023, pp. 527–532. doi: 10.1109/DeSE58274.2023.10099808.

[18] F. H. Awad and M. M. Hamad, “Improved k-Means Clustering Algorithm for Big Data Based on Distributed SmartphoneNeural Engine Processor,” 2022. doi: 10.3390/electronics11060883.

[19] A. Ahmed, K. Saleem, O. Khalid, J. Gao, and U. Rashid, “Trust-aware denoising autoencoder with spatial-temporal activity for cross-domain personalized recommendations,” *Neurocomputing*, vol. 511, pp. 477–494, 2022, doi: 10.1016/j.neucom.2022.09.023.

[20] F. H. Awad, M. M. Hamad, and L. Alzubaidi, “Robust Classification and Detection of Big Medical Data Using Advanced Parallel K-Means Clustering, YOLOv4, and Logistic Regression,” 2023. doi: 10.3390/life13030691.

[21] H. Wen, W. Guo, and X. Li, “A novel deep clustering network using multi-representation autoencoder and adversarial learning for large cross-domain fault diagnosis of rolling bearings,” *Expert Syst. Appl.*, vol. 225, 2023, doi: 10.1016/j.eswa.2023.120066.

[22] J. Wang, M. Gao, Z. Wang, C. Lin, W. Zhou, and J. Wen, “Ada: Adversarial learning based data augmentation for malicious users detection,” *Appl. Soft Comput.*, vol. 117, p. 108414, 2022, doi: 10.1016/j.asoc.2022.108414.

[23] Y. Lai, Y. Zhu, W. Fan, X. Zhang, and K. Zhou, “Toward adversarially robust recommendation from adaptive fraudster detection,” *IEEE Trans. Inf. Forensics Secur.*, vol. 19, pp. 907–919, 2023.

[24] A. S. Albahri *et al.*, “Trust and explainability in robotic hand control via adversarial multiple machine learning models with EEG sensor data fusion: A fuzzy decision-making solution,” *Comput. Biol. Med.*, vol. 196, p. 110922, 2025, doi: 10.1016/j.compbiomed.2025.110922.

[25] W. Fan *et al.*, “Adversarial Attacks for Black-Box Recommender Systems via Copying Transferable Cross-Domain User Profiles,” *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12415–12429, 2023, doi: 10.1109/TKDE.2023.3272652.

[26] F. Hazzaa, M. M. Hasan, A. Qashou, and S. Yousef, “A new lightweight cryptosystem for IoT in smart city environments,” *Mesopotamian J. CyberSecurity*, vol. 4, no. 3, pp. 46–58, 2024.

[27] Z. Shen, X. Guo, B. Feng, H. Cheng, S. Ni, and H. Dong, “Adversarial learning based residual variational graph normalized autoencoder for network representation,” *Inf. Sci. (Ny).*, vol. 640, Sep. 2023, doi: 10.1016/j.ins.2023.119055.

[28] A. E. Cinà, A. Torcinovich, and M. Pelillo, “A black-Box adversarial attack for poisoning clustering,” *Pattern Recognit.*, vol. 122, 2022, doi: 10.1016/j.patcog.2021.108306.

[29] L. Yang, S. X. Yang, Y. Li, Y. Lu, and T. Guo, “Generative Adversarial Learning for Trusted and Secure Clustering in Industrial Wireless Sensor Networks,” 2023. doi: 10.1109/TIE.2022.3212378.

[30] F. A. Al-Ibraheemi *et al.*, “Intrusion Detection in Software-Defined Networks: Leveraging Deep Reinforcement

Learning with Graph Convolutional Networks for Resilient Infrastructure.," *Fusion Pract. Appl.*, vol. 15, no. 1, 2024.

[31] Y. Wang, Y. Liu, Q. Wang, and C. Wang, "ClusterPoison: Poisoning Attacks on Recommender Systems with Limited Fake Users," *IEEE Commun. Mag.*, vol. 62, no. 11, pp. 136–142, 2024, doi: 10.1109/MCOM.001.2300558.

[32] L. Li, J. Xiahou, F. Lin, and S. Su, "DistVAE: Distributed Variational Autoencoder for sequential recommendation," *Knowledge-Based Syst.*, vol. 264, p. 110313, 2023, doi: 10.1016/j.knosys.2023.110313.

[33] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and P. Group, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *Int. J. Surg.*, vol. 8, no. 5, pp. 336–341, 2010.

[34] M. E. Alqaysi, A. S. Albahri, and R. A. Hamid, "Diagnosis-based hybridization of multimedical tests and sociodemographic characteristics of autism spectrum disorder using artificial intelligence and machine learning techniques: a systematic review," *Int. J. Telemed. Appl.*, vol. 2022, no. 1, p. 3551528, 2022.

[35] S. S. Joudar, R. A. Hamid, and I. A. Zahid, "Artificial intelligence-based approaches for improving the diagnosis, triage, and prioritization of autism spectrum disorder: a systematic review of current trends and open issues," *Artif. Intell. Rev.*, vol. 56, no. Suppl 1, pp. 53–117, 2023.

[36] H. A. Al-Tameemi *et al.*, "A Systematic Review of Metaverse Cybersecurity: Frameworks, Challenges, and Strategic Approaches in a Quantum-Driven Era," *Mesopotamian J. CyberSecurity*, vol. 5, no. 2, pp. 770–803, 2025, doi: 10.58496/MJCS/2025/045.

[37] J. Yu, L. Zhao, S. Yin, and M. Ivanovic, "News Recommendation Model Based on Encoder Graph Neural Network and Bat Optimization in Online Social Multimedia Art Education," *Comput. Sci. Inf. Syst.*, vol. 21, no. 3, pp. 989–1012, 2024, doi: 10.2298/CSIS231225025Y.

[38] J. Chen *et al.*, "N2VSCDNNR: A Local Recommender System Based on Node2vec and Rich Information Network," 2019. doi: 10.1109/TCSS.2019.2906181.

[39] R. Kashef, "Enhancing the role of large-scale recommendation systems in the IoT context," *IEEE Access*, vol. 8, pp. 178248–178257, 2020, doi: 10.1109/ACCESS.2020.3026310.

[40] M. Bazargani, S. H.Alizadeh, and B. Masoumi, "Group deep neural network approach in semantic recommendation system for movie recommendation in networks online," *Electron. Commer. Res.*, 2024, doi: 10.1007/s10660-024-09897-4.

[41] J. Chinnadurai *et al.*, "Enhancing online education recommendations through clustering-driven deep learning," *Biomed. Signal Process. Control*, vol. 97, 2024, doi: 10.1016/j.bspc.2024.106669.

[42] Y. Niu, Y. Su, S. Li, S. Wan, and X. Cao, "Deep adversarial autoencoder recommendation algorithm based on group influence," *Inf. Fusion*, vol. 100, p. 101903, 2023, doi: 10.1016/j.inffus.2023.101903.

[43] N. Khouibiri, Y. Farhaoui, and A. El Allaoui, "Design and Analysis of a Recommendation System Based on Collaborative Filtering Techniques for Big Data," *Intell. Converg. Networks*, vol. 4, no. 4, pp. 296–304, 2023, doi: 10.23919/ICN.2023.0024.

[44] Y. Zhang and X. Gu, "Enhancing user and item representation with collaborative signals for KG-based recommendation," *Neural Comput. Appl.*, vol. 36, no. 12, pp. 6681–6699, 2024, doi: 10.1007/s00521-024-09419-1.

[45] A. Chhabra, A. Roy, and P. Mohapatra, "Suspicion-free adversarial attacks on clustering algorithms," 2020, *Univ Calif Davis, tDept Comp Sci, Davis, CA 95616 USA*. doi: 10.1609/aaai.v34i04.5770.

[46] C. Zhang and Z. Tang, "Novel poisoning attacks for clustering methods via robust feature generation," *Neurocomputing*, vol. 598, 2024, doi: 10.1016/j.neucom.2024.127925.

[47] H. Xia, S. Shao, C. Hu, R. Zhang, T. Qiu, and F. Xiao, "Robust Clustering Model Based on Attention Mechanism and Graph Convolutional Network," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 5, pp. 5203–5215, 2023, doi: 10.1109/TKDE.2022.3150300.

[48] W. Yang, M. Wang, C. Tang, X. Zheng, X. Liu, and K. He, "Trustworthy multi-view clustering via alternating generative adversarial representation learning and fusion," *Inf. Fusion*, vol. 107, 2024, doi: 10.1016/j.inffus.2024.102323.

[49] Q. Lv, N. Yang, A. Slowik, J. Lv, and A. Yousefpour, "Market behavior-oriented deep learning-based secure data analysis in smart cities," *Comput. Electr. Eng.*, vol. 108, 2023, doi: 10.1016/j.compeleceng.2023.108722.

[50] X. Ye, J. Zhao, Y. Chen, and L. J. Guo, "Bayesian Adversarial Spectral Clustering with Unknown Cluster Number," *IEEE Trans. Image Process.*, vol. 29, pp. 8506–8518, 2020, doi: 10.1109/TIP.2020.3016491.

[51] C. Pimsarn, T. Boongoen, N. Iam-On, N. Naik, and L. Yang, "Strengthening intrusion detection system for adversarial attacks: improved handling of imbalance classification problem," 2022. doi: 10.1007/s40747-022-00739-0.

[52] H. Shirazi, B. Bezawada, I. Ray, and C. Anderson, "Directed adversarial sampling attacks on phishing detection," *J. Comput. Secur.*, vol. 29, no. 1, pp. 1–23, 2021, doi: 10.3233/JCS-191411.

[53] P. Tatongjai, T. Boongoen, N. Iam-On, N. Naik, and L. Yang, "Classification of Adversarial Attacks Using Ensemble Clustering Approach," *Comput. Mater. Contin.*, vol. 74, no. 2, pp. 2479–2498, 2023, doi: 10.32604/cmc.2023.024858.

[54] W. Tang, B. Hui, L. Tian, G. Luo, Z. He, and Z. Cai, "Learning disentangled user representation with multi-view information fusion on social networks," *Inf. Fusion*, vol. 74, pp. 77–86, 2021, doi: 10.1016/j.inffus.2021.03.011.

[55] G. Delorme, Y. Xu, S. Lathuilière, R. Horaud, and X. Alameda-Pineda, "CANU-ReID: a conditional adversarial network for unsupervised person re-identification," in *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, 2021, pp. 4428–4435.

[56] C. Sandoval, E. Pirogova, and M. Lech, "Adversarial learning approach to unsupervised labeling of fine art paintings," 2021. doi: 10.1109/ACCESS.2021.3086476.

[57] S. Rass, S. Konig, S. Ahmad, and M. Goman, "Metricizing the Euclidean Space Toward Desired Distance Relations in Point Clouds," *IEEE Trans. Inf. Forensics Secur.*, vol. 19, pp. 7304–7319, 2024, doi: 10.1109/TIFS.2024.3420246.

[58] D. He *et al.*, "Adversarial representation mechanism learning for network embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 2, pp. 1200–1213, 2021.

[59] H. Ma *et al.*, "Stealthy attack on graph recommendation system," *Expert Syst. Appl.*, vol. 255, p. 124476, 2024, doi: 10.1016/j.eswa.2024.124476.

[60] Q. Wang, C. Wu, D. Lian, and E. Chen, "Securing recommender system via cooperative training," *World Wide Web*, vol. 26, no. 6, pp. 3915–3943, 2023, doi: 10.1007/s11280-023-01214-7.

[61] T. Baker, T. Li, J. Jia, B. Zhang, C. Tan, and A. Y. Zomaya, "Poison-Tolerant Collaborative Filtering Against Poisoning Attacks on Recommender Systems," *IEEE Trans. Dependable Secur. Comput.*, vol. 21, no. 5, pp. 4589–4599, 2024, doi: 10.1109/TDSC.2024.3354462.

[62] C. Wu, D. Lian, Y. Ge, Z. Zhu, and E. Chen, "Influence-Driven Data Poisoning for Robust Recommender Systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 11915–11931, 2023, doi: 10.1109/TPAMI.2023.3274759.

[63] Q. Zhou, K. Li, and L. Duan, "Recommendation attack detection based on improved Meta Pseudo Labels," *Knowledge-Based Syst.*, vol. 279, 2023, doi: 10.1016/j.knosys.2023.110931.

[64] H. Cai, J. Ren, J. Zhao, S. Yuan, and J. Meng, "KC-GCN: A Semi-Supervised Detection Model against Various Group Shilling Attacks in Recommender Systems," *Wirel. Commun. Mob. Comput.*, vol. 2023, 2023, doi: 10.1155/2023/2854874.

[65] Y. Du, M. Fang, J. Yi, C. Xu, J. Cheng, and D. Tao, "Enhancing the robustness of neural collaborative filtering systems under malicious attacks," *IEEE Trans. Multimed.*, vol. 21, no. 3, pp. 555–565, 2018.

[66] H. Liu, L. Guo, P. Li, P. Zhao, and X. Wu, *Collaborative filtering with a deep adversarial and attention network for cross-domain recommendation*, vol. 565. 2021. doi: 10.1016/j.ins.2021.02.009.

[67] Z. W. Wu, C. T. Chen, and S. H. Huang, "Poisoning attacks against knowledge graph-based recommendation systems using deep reinforcement learning," *Neural Comput. Appl.*, vol. 34, no. 4, pp. 3097–3115, 2022, doi: 10.1007/s00521-021-06573-8.

[68] J. Zhao, J. Fang, P. Chao, B. Ning, and R. Zhang, "GC-TripRec: Graph contextualized generative network with adversarial learning for trip recommendation," *World Wide Web*, vol. 26, no. 5, pp. 2291–2310, 2023, doi: 10.1007/s11280-022-01127-x.

[69] O. Papakyriakopoulos, J. C. M. Serrano, and S. Hegelich, "Political communication on social media: A tale of hyperactive users and bias in recommender systems," *Online Soc. Networks Media*, vol. 15, p. 100058, 2020, doi: 10.1016/j.osnem.2019.100058.

[70] Y. Zhao, K. Wang, G. Guo, and X. Wang, “Learning compact yet accurate Generative Adversarial Networks for recommender systems,” *Knowledge-Based Syst.*, vol. 257, 2022, doi: 10.1016/j.knosys.2022.109900.

[71] X. Dai, Z. Wang, J. Xie, T. Yu, and J. C. S. Lui, “Online Learning and Detecting Corrupted Users for Conversational Recommendation Systems,” *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 12, pp. 8939–8953, 2024, doi: 10.1109/TKDE.2024.3448250.

[72] F. Zhang, P. P. K. Chan, Z. M. He, and D. S. Yeung, “Unsupervised contaminated user profile identification against shilling attack in recommender system,” *Intell. Data Anal.*, vol. 1, no. 1, 2024, doi: 10.3233/IDA-230575.