International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Textual Online Content Recommendation Towards Accountability: A Thorough Disquisition of Recent Advances, Problems, and Potential Scope Anatomizing Obscure Features.

[1]Aanal Raval, [2]Komal R. Borisagar

Submitted: 06/05/2024 **Revised**: 19/06/2024 **Accepted**: 26/06/2024

Abstract: The vast expanse of online information often overwhelms users, leaving them unable to navigate through the wealth of available data. Recommender systems have emerged as a remedy to aid users by suggesting relevant items based on their preferences. However, many challenges persisted, hampering these systems. This paper conducts a thorough literature review, delving into various approaches aimed at mitigating issues in recommender systems. It explores diverse similarity measures, strategies within different recommender systems, and the use of multimodal data. Additionally, the study evaluates the effectiveness of these approaches in tackling the challenges focusing on evaluation metric utilities. The findings underscore the significance of addressing sparsity, providing valuable insights for researchers and practitioners to develop more robust recommender systems, ultimately enhancing recommendation accuracy and efficiency.

Keywords: *practitioners, ultimately, recommendation, mitigating,*

1. Introduction:

The digital age has ushered in an abundance of online publications, ranging from research articles to legal and medical documents. However, sifting through this vast array of literature to find the right piece can be a daunting task for researchers, legal professionals dealing with judgments and laws, as well as individuals seeking job matches. While academic search engines and digital repositories exist to aid in this search, users often struggle to articulate precise keywords and filter through the multitude of results effectively, especially newcomers. Document recommendation systems offer a solution by assisting users in identifying relevant articles from extensive repositories, addressing the challenges of information overload and search precision.

Several existing surveys delve into various aspects of recommendation systems, with a focus on article citations, relevant article retrieval, or news story suggestions. Some studies, such as [56], center on paper recommendations, considering content context and examining citation behavior. Additionally, there are works that delve into news recommendation challenges [29], highlighting the significance of content quality and its impact on users, which extends to readers of diverse content types. However, these studies often overlook the influence of images as a feature, along with essential considerations like privacy and decentralized computations. This review aims to fill this gap by exploring different document recommendation systems and shedding light on unexplored avenues. By analyzing the current state-of-the-art techniques and findings, this survey aims to aid in the advancement of innovative developments in this field.

[1]*Research Scholar Gujarat Technological University*
aanalraval@gtu.edu.in Orchid Id: 0000-0003-3848-0882
[2]*Associate Professor Gujarat Technological University*
asso_komal_borisagar@gtu.edu.in Orchid Id: 0000-0002-1047-6091

2. Document Recommendation Aspects:

A vast array of reading materials, spanning various fields such as technical, non-technical, legal, medical documents, news articles, and research papers, are published regularly. Navigating through this extensive, continuously expanding, and unstructured data landscape depends significantly on factors like the material's recency and popularity. Typically, readers can access this content without signing in or creating profiles, leading to limited feedback and reliance on consumption behavior analysis through click logs, session tracking, or browser cookies. This prompted the development of user profiles, primarily to facilitate recommendations by using historical data. However, challenges emerge when new users or items enter the system, presenting scenarios with minimal or no available data—a situation known as the cold start problem.

Besides the cold start challenge, user preferences exhibit temporal variations, comprising both short-term and long-term inclinations. Furthermore, item churn, or the changing availability of content, significantly impacts recommendations. In the realm of document or reading recommendations, there's a risk of over-specialization and filter bubbles, where only specific content is suggested, potentially limiting exposure to diverse and current suggestions. This has tendency to narrow users' perspectives and choices by favoring certain content, despite lesser popularity, might highlight the importance of handling long-tail scenarios to ensure a broader range of recommended content.

3. Methodology:

The study aims to analyze the trends in recommender system research, which spans multiple disciplines. Papers on this topic are found in diverse journals like computer science, management, marketing, and information technology, making it challenging to confine within a single discipline. To conduct this literature review databases of research andvarious electronic journals like IEEE/IEE Library, ACM Portal, Science Direct and Google Scholars are being considered to ensure comprehensive coverage. Based on four descriptors: "Textual Recommender systems", "Document Recommendation", "Article Recommendation", "Bulk text Recommendation", this review has been conducted.

3.1 Quality analysis:

To fortify the credibility of our study, immediately following the comprehensive reading of all included papers, each publication within the final set underwent a rigorous assessment to evaluate its quality. More emphasis was laid on Investigation of sparsity and long tail (RQ1), consideration analysis of multimodal data (RQ2), Analysis of cold start (RQ3) and Explainability of Recommendations (RQ4). While evaluation metric utility (RQ5) received comparatively less attention.

3.2 Criteria for inclusion or exclusion:

We have analysed the papers dealing with textual bulk. Sometimes concerning any research paper or with any textual bulk case, image and link or citation or reference too contribute a lot to the context. So here in this study, we consider all types of textual bulk either dense or sparse, images, links or references, social relations, trust, demographics and physiology, excluding just video and audio data.

4. Review of research in Recommendations dealing with textual bulk:

4.1 Literature review:

In this section, we outline the current research and analyze it across various dimensions, categorizing document recommendation tasks into different subcategories. The accompanying figure illustrates the diverse application categories (Figure-1).

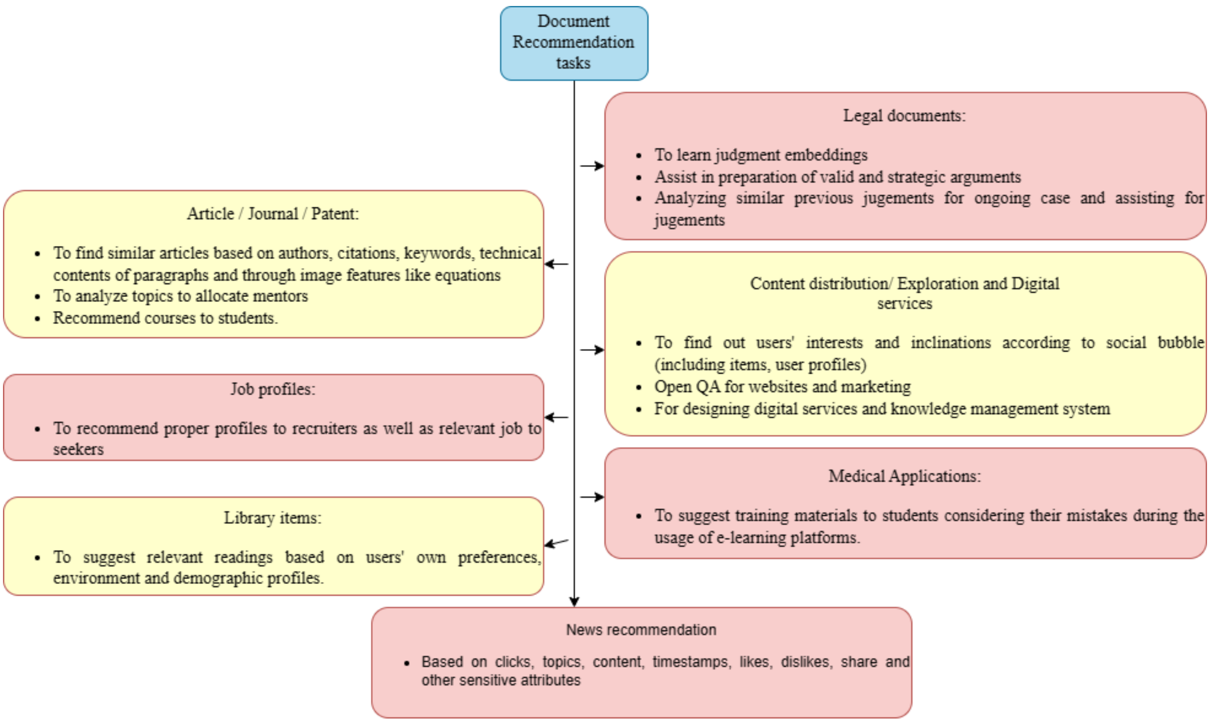

Fig-1 Categories of Document Recommendation applications along key tasks or goals to be done [115]

Legal Document Recommendations: Utilizing pre-learned word embeddings specific to the legal domain, [1] introduced the P-LDRS framework, incorporating legal semantic knowledge. This framework facilitated the learning of Doc2vec embeddings, including judgment embeddings. To tackle scalability, the framework employed distributed computing across MapReduce and Spark-like node clusters. Results showcased the framework's superiority over traditional Doc2vec, achieving an impressive 88% accuracy. In a similar vein, [2] proposed a methodology aimed at eliminating noisy data and reducing corpus size by preserving relevant dictionary words while discarding extraneous elements. This approach aimed to enhance the efficiency of Doc2vec models. Addressing the exponential growth of judgments over time [73], relevant judgments were identified from each cluster using pairwise similarity scores.

Journal / Technical or Non-technical Article/ Patent / citation recommendation: Several studies have explored diverse methods for document recommendation systems. [3] developed a recommender system based on ThaiJO using tf-idf and cosine similarity scores, while [4] incorporated word2vec alongside tf-idf for vector weighting. [6] introduced a two-stage deep learning model utilizing patent text and citation information, outperforming existing models. [7] leveraged doc2vec and XGBoost for context-based article recommendations, achieving an 84.24% accuracy. [8] proposed a transfer learning-based model using RoBERTa, reaching 94.96% accuracy in top 10 recommendations. Other approaches included [9]'s multitask learning for metadata prediction and [10]'s hierarchical attention network for missing reference recommendations. [11] and [12] employed text clustering and alignment techniques, respectively, for document recommendations. Similarly, [16] used Latent Dirichlet Allocation (LDA) for semantic structure extraction, while [19] employed search intent and text mining techniques for user behavior analysis. Methods like tf-idf with K-means ([20]) and LDA with distance measures ([22] and [23]) were also explored for article recommendation. Additionally, [27] introduced features like Keyword Diversity Measurement and Citation Analysis Over Time, while [28] utilized LDA and Latent Semantic Analysis (LSA). Furthermore, [29] constructed a heterogeneous relation graph for personalized recommendations, while [62] focused on personalized learning through heterogeneous networks. [64] proposed one-shot learning and attention-based approaches but observed limitations in personalized settings, while [65] utilized a citation network for paper recommendations. Lastly, [70] addressed the cold start problem by reinforcing co-author relations with textual information. This summary highlights various methodologies such as tf-idf, word embeddings, deep learning models, clustering, and network-based approaches employed in document recommendation systems across different studies.

Content distribution/ Digital Exploration recommendations: The information provided outlines diverse approaches within document recommendation systems. For instance, [15] introduced SEAN (social explorative attention network), leveraging attention mechanisms to encode documents into user-dependent contextual vectors. Variants like SEAN-END2END and SEAN-KEYWORD were designed to cater to broader user scopes and new users with limited history. In the domain of knowledge management systems for sales, [21] developed similarity-based recommendation models for proposals by extracting industrial context features and applying word embedding models. Similarly, [25] utilized LSTM for preprocessing and graph embeddings for topic and content recommendations, while [26] used various recommendation techniques—content-based, collaborative filtering, and complementarity-based—for website QA document modeling. For quotation tasks, [14] employed BERT-based models for content-based recommendations by ranking and scoring paragraph contents, integrating open-question answering for enhanced recommendations. Additionally, [67] proposed a deep learning-based approach with bidirectional encoders and sequential recommendation models for e-commerce platforms, while [68] focused on document ranking considering query signals and user preferences. In the realm of comment appropriateness on Stack Overflow, [71] suggested feature-based evaluation using NLP techniques such as cosine and Jaccard similarity, and random forest clustering to assess the relatedness among comments. These features encompassed various factors like text similarity, semantic features, and temporal aspects to gauge **pairwise connectedness and cluster creation.**

Job profile recommendations: In the realm of job recommendations, [17] employed an approach centered on implicit skills extracted through NLP techniques. They utilized document similarity matching and projected these skills into a semantic space using Doc2vec. Implicit skills, not explicitly mentioned in resumes but relevant to specific roles, industries, or geographies, were considered. The goal was to match candidates' CVs to job descriptions by considering various skills. Meanwhile, [63] utilized a range of NLP techniques including NLP-token count matrix, cosine similarity percentages, phrase matching, and Spacy for feature creation and matching in the context of calculating job satisfaction and retention parameters. Linear and ridge regression methods were employed, exploring several techniques such as Naïve Bayes, Support Vector Classifier, K-nearest neighbor, decision trees, random forests, and Multilayer perceptron for parameter calculation.

Library items and other Academic applications: In the domain of recommendation systems, various innovative approaches have been explored. [5] introduced the Knowledge-based Environment architecture for Personalized Learning (KEPLAIR) using Artificial Intelligence Recommendation. This framework comprises three core modules: the Learning Manager, supervising interactions and recommending learning materials; the Social Manager, logging user communications and behaviors; and the Harvesting Manager, storing evolving metadata of learning objects. KEPLAIR's knowledge base integrates user interests, demographics, personal preferences, and learning environments. Experimental results showed promising prediction rates when prior user interest information was available. [13] focused on document collection, employing graph-based ranking within content-based recommender systems, deriving indices from abstracts. Meanwhile, [18] developed a methodology for comparing topic proposals to allocate mentors, utilizing Rapid Automatic Keyword Extraction and Levenshtein distance similarity for keyword generation and document matching, respectively.Additionally, studies like [59] and [60] observed students' examination performance and advised on course selection based on subject-specific performance. Lastly, [72] addressed the issue of handling discussions in Massive Open Online Courses (MOOCs), proposing a system to analyze chat conversations. This system recommends appropriate learning materials considering chat sentiment, intent, and the clarity or confusion of concepts discussed, providing a solution for experts facing challenges in addressing every individual in the forums.

Medical field Usages: The exploration of recommendation systems for diverse applications continues to evolve. In e-learning platforms, [24] experimented with Learning to Rank methods like Lambda Rank, MART, ListNet, RankNet, LambdaMART, RankBoost, Random Forest (RF), and Coordinate Ascent. Their study highlighted that RF achieved the highest Normalized Discounted Cumulative Gain (NDCG) of 0.89 in scenarios without new documents. For medical recommendations, [58] utilized fuzzy inference and decision tree rules to suggest medicines and preventive measures for diabetes, emphasizing natural care, exercises, nutrition, and medications. Similarly, [62] explored ensemble learning methods, traditional techniques, embeddings, and clustering like SMOTE and t-SNE for drug recommendations based on reviews. In the domain of online healthcare communities, [69] developed a system to aid by recommending necessary steps based on patients' needs, QA content, and doctors' backgrounds. They utilized embeddings and an attention-based mechanism to capture interactions, applying a Gated Recurrent Unit (GRU) to match scores and generate a ranking list. These studies showcase varied approaches in recommendation systems across e-learning, medical advice, and online healthcare communities.

News recommendations: Research in news recommendation systems is diverse and involves numerous approaches. [29] delved into traditional methods such as collaborative filtering (CF), content-based filtering (CBF), and hybrid models, including factorization techniques like matrix factorization (MF), Non-negative matrix factorization (NMF), Tensor factorization, Probabilistic matrix factorization (PMF), Bayesian personalized ranking (BPR), Generalized linear modeling (GLM), and their neural extensions. They also discussed various libraries and platforms like Apache Mahout, Idomaar, StreamingRec, CLEF NEWSREEL, Open Recommendation Platform (ORP), and Hugging Face for news recommendation development. Moreover, the study explored DL-based models like Multi-layer perceptron (MLP), Autoencoder (AE), Convolutional neural network (CNN), Recurrent neural network (RNN), attention modules, Graph neural network (GNN), Transformers, and Reinforcement learning (RL) for recommendation purposes. [47] surveyed methods encompassing entity-centric, path-based, and neural-based approaches, utilizing latent representations and knowledge graphs. In specific experiments, [30] explored context-based features like title, popularity, freshness, and clicks across various existing models, while [44] used hierarchical attention for subtopic interest modeling. [45] employed an attention network for user modeling, considering the relatedness of clicked news. Furthermore, [48] deployed a hypergraph with affinity matrices and ranking based on timestamps, stakeholder weights, and coverage weights of authors. To counter over-specialization and under-representation, they used author popularity, constructing a graph to scale weights as per user preferences. Their approach aimed at diversity by considering relevance scores for unexplored topics and keywords in the recommendations.

Recommendations in Social network: [101] treated missing entities from “exposure to customer” perspective and revealed the ratings with causality. For this purpose they proposed De-Bias Network Confounding in Recommendation for analysis of co-founders in casual inference, where DENC addresses Missing Not At Random (MNAR) data by analyzing inherent factors (like hidden user or item traits) and auxiliary networks. Its exposure model manages social network influences while preserving observed exposure data. Additionally, DENC incorporates a deconfounding model through balanced representation learning to maintain essential user and item features, improving rating prediction accuracy. [102] Used raw data from sensors, then extracted geographical information through POI, semantic information through CBOW and recognize category along with temporal feature space. Next a Geographical-Spatiotemporal Gated Recurrent Unit Network (GSGRUN) selects relevant check-in activities to alter user preferences. [103] considered the issue of overlapping social trust relationships in overlapping communities for evaluation of modifications and achieving consensus in Social network group decision making (SN-GDM). [104] Proposed reinforcement learning based Policy learning for Social Group Recommendation (IP-SGR) method where co-interacting items are captured and aggregated through gate filtering, followed by capturing and representation of group by inter-group attention network and items by inter-item prototype learning model and finally to obtain a reward for social group. [105] Developed a methodology FedPOIRec for aggregating user friends' preferences by enabling knowledge transfer among friends through federated learning and homomorphic encryption. [111] Deployed federated learning and graph neural network assisted with pseudo-labeling technique in item sampling for privacy protection and enhancement of training data. [106] Using tags and tweets as features deployed k nearest neighbour for social circle recommendation. [107] integrated Graph convolutions with representation learning to calculate social trust using influence and trust among friends. [109] Targeted Short and long term preferences for social diversity taking Embeddings of behaviour and knowledge features along with semantic associations as features considering network heterogeneous relationships for knowledge graphs. [108] Modelled user-item interactions with Social light graph convolution for Social recommendation. [110] Studied determinants affecting user trust by Fedrated Social recommendation with Graph neural network (FeSoG) for Social Recommendation.

4.2 Description of Datasets:

Given that the focus of both recommendation studies revolves around voluminous text, our analysis centers on textual document data. In this discussion, we explore various datasets sourced from public repositories, synthesized collections, or obtained through web crawling. These datasets encompass a wide array of text-based information for examination and utilization in recommendation system research.

The FullTextPeerRead dataset is context-aware, housing both cited references and paper metadata. Similar in nature, the ANN dataset also includes attributes such as target and source IDs, along with additional citation text. In contrast, the ACL-200 dataset offers papers embedded within contexts. RefSeer, released under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License, comprises three primary tables: 'paper' (featuring IDs and their respective clusters), 'citations' (including citation IDs, clusters, and paper IDs), and 'citation index' (with citation IDs and contextual information) [76]. For expansive scholarly content, the arXiv dataset grants access to approximately 1.7 million open-access articles spanning computer engineering and diverse fields like statistics, mathematics, economics, quantitative biology, and electrical engineering. With features like titles, categories, authors, abstracts, and full text, this dataset serves various purposes, from paper recommendation and trend analysis to co-citation networks, semantic search interfaces, category prediction, and constructing knowledge graphs [77]. Additionally, the HepTh dataset is a graph network dataset illustrating relationships among authors and their respective citations.

The Aminer dataset, a non-commercial resource, establishes connections between researchers alongside statistical correlations. Spanning 130,000,000 researchers and over 265 million publications, this dataset serves multiple purposes, aiding in social network analysis. It facilitates tasks such as topic modeling, evaluating academic performance, expert identification, review recommendations, course and geographic searches, and association searches [79]. Derived from DBLP, ACM, and Microsoft Academic Graph, the DBLP dataset is a citation network dataset housing more than 632,752 citations and 629,814 papers. Each paper within this dataset includes title, abstract, authors, venue, and publication year. Beyond supporting topic modeling and network clustering, it uniquely explores mutual influence within the citation network, identifying highly influential papers [80]. RARD (Related-Article Recommendation Dataset) originates from Mr. DLib and the digital library Sowiport. It encompasses recommendation approaches, feature types, extraction sources, time logs, and click logs [82]. Notably, its successor, the second version, boasts a 187% increase in features, 64% more recommendations, 50% additional clicks, and 140% more documents compared to its predecessor.

The Drug Review dataset compiles reviews on specific drugs alongside customers' ratings. In the legal domain, the JRC-Acquis stands as a multilingual corpus sourced from the European Union [78]. Similar Skills and Skill2vec datasets consist of over 1,400,000 job descriptions, extracting skills and establishing contextual similarity among them, specifically relating to job descriptions [81]. The Wikipedia Graph and Related Entity Recommendation Dataset features a normalized and detailed Wikipedia hyperlink entity graph, encompassing multiple languages and distinct features. It includes query entities paired with relevant entities, serving as ground truth data for evaluation purposes. The entity embeddings are trained via lg2vec. Moreover, the Yahoo Search Query Log To Entities dataset serves as a benchmark for entity linking. It helps in associating search queries with relevant entities.

Lastly, the Yahoo Answers Novelty Based Answer Ranking dataset focuses on medical content. It annotates relevant context and answers for target questions, serving as a mapping of textual assertions.

The Roularta dataset, sourced from a Belgian multimedia group, encompasses weekly/monthly news magazines spanning business trends, sports, women's interests, lifestyle, local news, medical insights, and other professional domains. It includes content summaries, titles, and interlinked information from various articles. This dataset utilizes IPTC (International Press Telecommunications Council) tags, facilitating news exchange between different agencies. Notably, author popularity within user interactions exhibits skewed distributions. The dataset employs CNN for text embedding generation and feature extraction, shedding light on both prominent and less visible authors. The Plista dataset amalgamates news data from 13 different portals, offering statistical analyses involving Click-Through Rates (CTR), publisher activity, updates, and impressions. It reveals biases in popularity and sparsity, providing statistical summaries such as sums, means, and standard deviations of pertinent features [50]. Yahoo Webscope serves as a reference library offering datasets for scientific experiments and academic use. It includes benchmark datasets like the Front Page Today Module User Click Log Dataset, Yahoo News Video dataset, and the Yahoo News Annotated Comments Corpus. Moreover, the Yahoo News Ranked Multi-label Corpus facilitates multi-label tagging and ranking tasks, providing actual text from various sources like Newsroom, Yahoo News, and Tumblr for researchers to customize their features. The Yahoo! News Sessions Content dataset offers session-wise user information, encompassing clicks, article counts, tokens, publication details, timestamps, and entities. Notably, this dataset anonymizes user and news data to ensure user identity remains undisclosed.

The Adressa dataset showcases skewed distributions in author popularity within user interactions. Spanning 11,207 articles and 561,733 users, this dataset unravels connections between news items and anonymized users. It encompasses diverse information such as subscriber status, session identification, click counts, reading times, titles, authors, categories, keywords, entities, and article bodies [49]. Hacker News presents a dataset comprising stories and comments from the Hacker News platform. This dataset includes story IDs, author details, timestamps, and popularity metrics. Released under the MIT License, it serves as a platform for news sharing, IT project demonstrations, job postings, inquiries, and story commentaries. Additionally, BuzzFeed News offers a dataset featuring news stories, entertainment content, and insights into fake news patterns within social media and news platforms.

The Microsoft news dataset (MIND) stands as a significant benchmark in the domain of news recommendation. Comprising approximately 160,000 news articles and over 15 million impression logs from one million users, this extensive dataset is pivotal in the field. Each log includes a history of click behaviors, encompassing both click events and non-clicked events—pertinent features for each article span title, body, abstract, category, and entities. To safeguard user privacy, user behavior logs are anonymized. Regarding fake news datasets, the FakeNewsAMT dataset emerges from a combination of crowdsourced and manual annotations, covering diverse domains like education, technology, sports, business, entertainment, and politics. Meanwhile, the Celebrity dataset centers on stories about socialites, politicians, singers, and actors. In the Celebrity dataset, legitimate news stories are sourced from fashion and entertainment sections, whereas the fake news category includes online gossip stories.

Additionally, datasets such as the Global Database of Events, Language, and Tone (GDELT), Reuters Corpora, 20 Newsgroups, and Fast.ai are widely employed for text categorization purposes.

4.3 Focus of Domain Research:

Recent works on textual recommender were analyzed, and observed that 4% of work from legal domain, 16% of work is technical/ non-technical articles or patent or citations related, 9% contributions are in content distributions or digital explorations, 3% focused on job recommendations, 9% of work contributed to library and other related academic applications, 6% wok focuses on medical materials, 26% of work forming one of majority, focused on news recommendations and major contributions are seen in recommendations for social networking.Figure-2 shows analysis of various recommender systems based on its application.

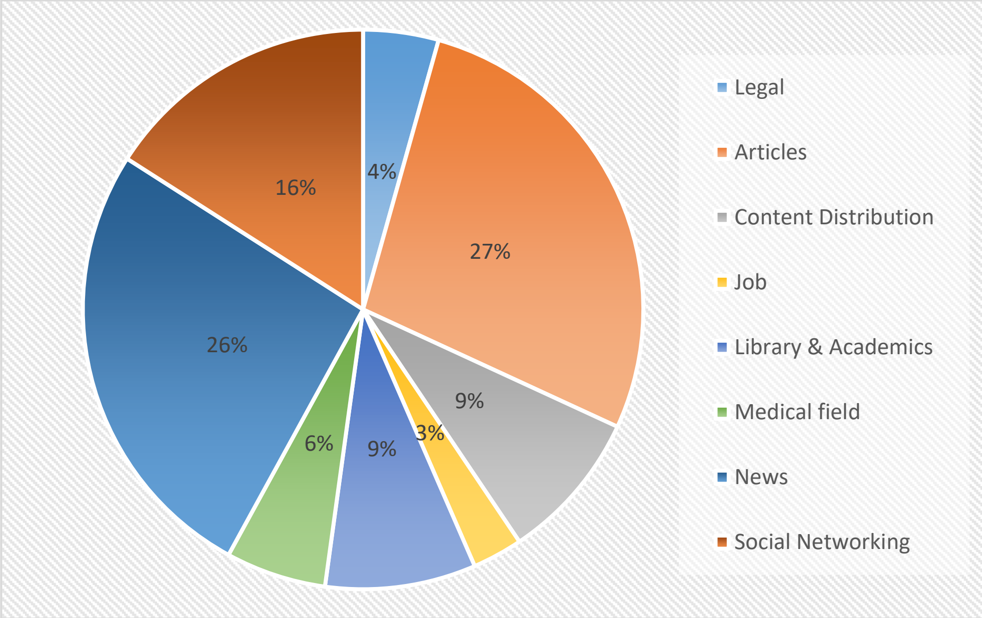

Fig-2 Application-based classification of recommendation system

Similarly, following Figure- 3 shows analysis of recent researches focusing the methodologies and techniques applied initially. Deployement of traditional recommender methods like collaborative filtering or content based were widely applied introducing more challenges. These challemges were tried to solve through various techniques from which clustering, ensemble methods and similarity score based methods showed up as comparatively better and Language models(LM), Encoders, Transformers, Attention-based methods, Graphical methods and Reinforcement learning focused techniques ended up as state of art methodologies in their respective tasks based on respective featured datasets.

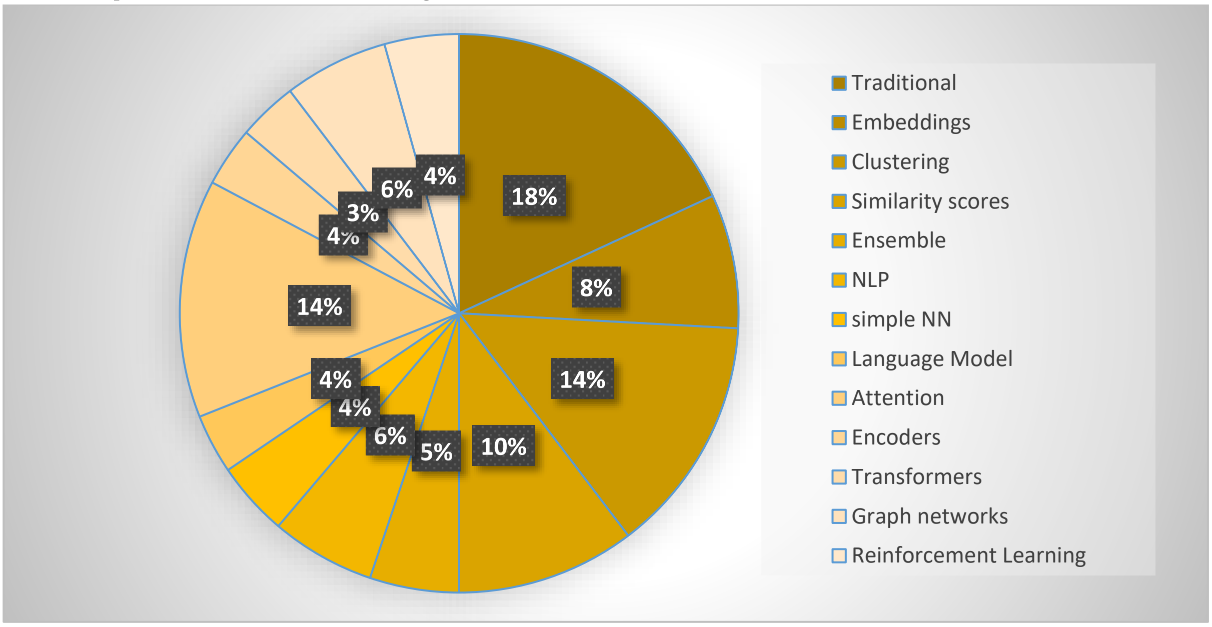

Fig-3 Classification of literature based on techniques and methodologies used.

5. Analysing the Performance of State of Art systems:

5.1 Findings:

A study of literature in the medical field reveals modelling of multiturn diagnostic QA documents while online interactions for recommending most common conditions, symptoms or questions from users prove effective in terms of utilizing medical resources, patients' economic resources [69], can save doctor's time in diagnosing of frequent diseases and can address with medical needs of patients that are unintentionally missed out [24]. [62] Recommends drugs based on sentiments. In general, patients always describe symptoms and suffering verbally. This shows the need to model the natural language to documents for further processing or even a chatbot-type interacting system can be introduced to ask some common questions. Sometimes apart from just QA, or symptoms description from patient, some medical tests reports, that can be in form of text document or image also becomes equally important, indicating mechanism for analysing image features and deriving the relevant need to be introduced. Sometimes scenario occurs when a particular drug receive negative sentiments but its long time effects are positive. Instead recommending drugs based on success rate, short term interventions, and side effects on normal person or with some medical complication or with different immune systems.

Exploring works for legal document recommendations, conclusions show citations [1] sometimes along with judgement texts are needed, which hasn't been considered. A kind of search engine or repository can be prepared [2] incorporating these embeddings for relevancy in n search. Besides, as decree always follow judgement, analysis can be performed on the pair. The case of "mistake of law" remains unexplored. For this case, RL can resolve better, where case of law of mistake document can be treated as irrelevant action and can be penalized. Thus learning the parameters or circumstances for same and suggest if there are any chances while referring old documents.

Examining the existing work, for rest of various applications, document semantic and reference recommendation lacks explainability. Furthermore retrieving relevant and concerned data from the vast voluminous pool as well as multiple origin still remains an issue. Additionally, most of proposals focus on single-single item recommendations, There is an immense need for single to multiple domain recommendations. Along with mentioned points, an important factor to be concerned includes tackling of heterogeneous data and multimodal estimations. Apart many other factors besides accuracy affects the model, which points towards the consideration of other related parameters too. Fairness of recommendation is an important aspect which is normally ignored [75]. Introducing hyper graph based method can provide considerable fairness [48]. Although, attention mechanism has been successfully deployed for this task showing great results, this methodology generally adds up the weights of respective hidden layers for creation of vector representation. This implies when attention mechanism is applied where session exits, where any user performs duplicated or frequent tasks, then this approach can generate similar estimations. Recommendation generally relies on the principal task of recommending along with the subordinate task of enhancing those estimations continuously [74]. Use of graph based methods can ensure the validity of data against noise and outliers to a great extent. Following Figure- 4 shows issues addressed.

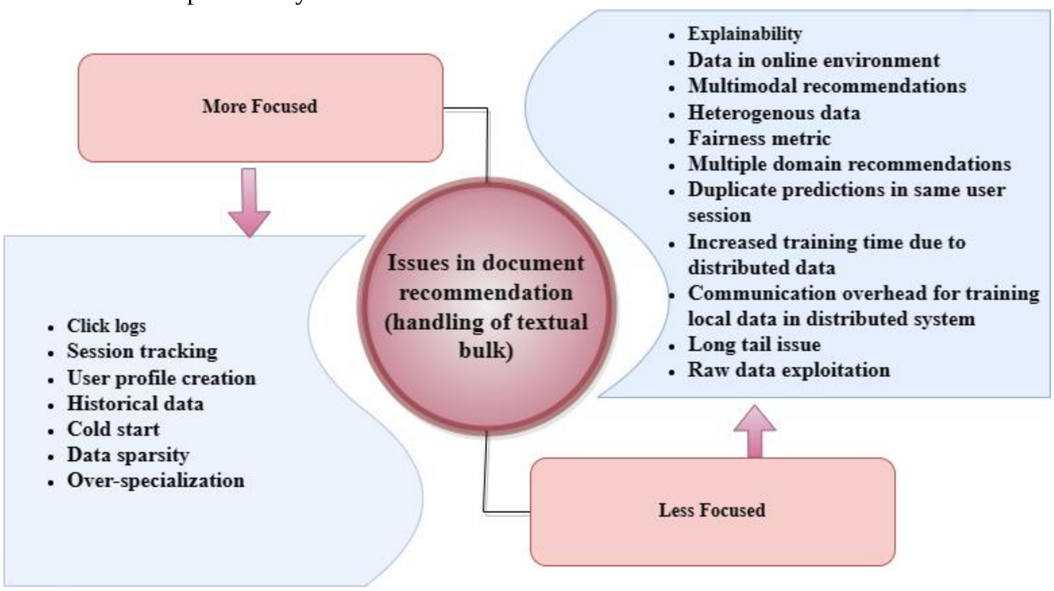

Fig-4 Issues or points in document recommendation (textual bulk handling) which are commonly explored as well as less explored and needs attention.

5.2 Analysis of Evaluation Metrics:

The motivation behind a recommender system is to predict the likelihood that a user will appreciate an unknown entity recommended to them. Understanding user preference patterns is crucial, as it determines if the suggested items align with their preferences. Previously, similarity-based metrics were utilized to assess the proximity of new items. Business-related metrics include click-through rates, adoption measures (effectiveness of suggestions), sales, revenue, distribution patterns, and user engagement. We classify evaluation metrics based on recommendation tasks. Following Figure-5 shows usage of metrics for various tasks and scenarios.

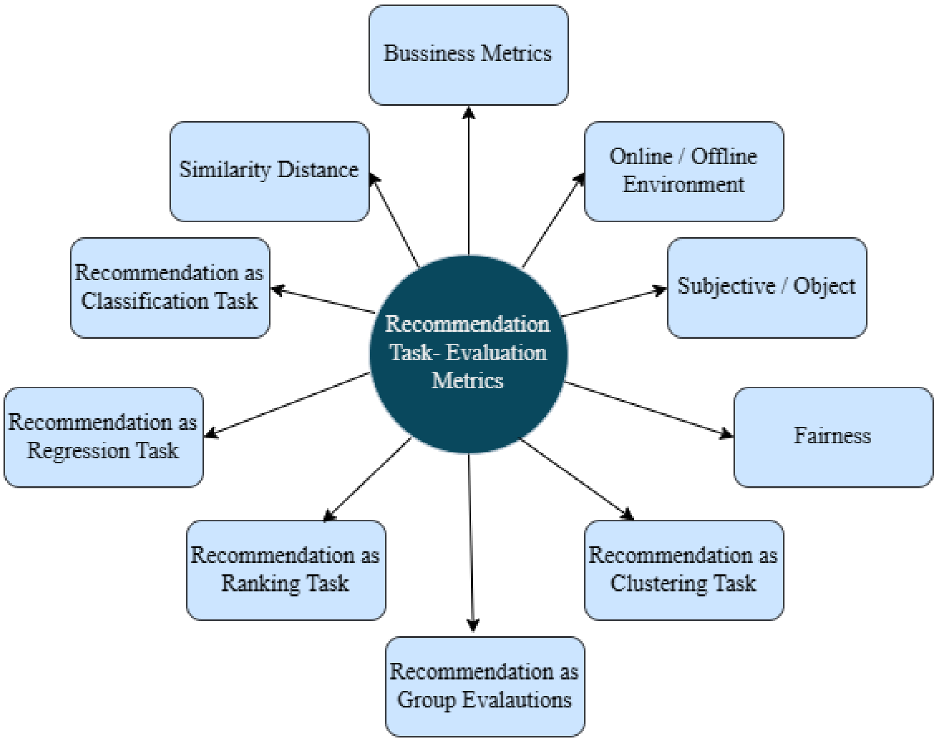

Fig 5 Usage of metrics for various tasks and scenarios[115]

*Classification tasks:*The classification-oriented approach to recommendation tasks relies on metrics such as accuracy, precision, F1 score, recall, and Area Under Curve (AUC). AUC, also known as the Receiver Operating Characteristic (ROC) score, gauges the likelihood of higher-ranking positive entities, presented as a probability curve plotting sensitivity versus 1-specificity. Sensitivity (or recall or true positive rate) indicates correctly identified positives, while specificity refers to the true negative rate, signifying accurately classified negatives. Conversely, 1-specificity portrays the false positive rate, indicating correctly categorized negative class proportions. Higher x-axis values in the AUC curve suggest more false positives (FP) than true negatives (TN), while higher y-axis values imply more true positives (TP) than false negatives (FN). Precision denotes the fraction of correctly classified true samples among all true samples. The Matthews Correlation Coefficient (MCC) measures the correlation between actual true classes and predicted labels. Unlike log loss, which quantifies differences in predicted and expected probability distributions, the Brier Score (BS) assesses the average difference in probabilities only, reflecting probability accuracy alongside Mean Squared Error (MSE).

*Regression tasks:*The regression-based approaches in recommendation systems employ metrics such as MAE, MSE, RMSE, PCC, RMSLE, R2, and adjusted R2. Mean Absolute Error (MAE) signifies the difference between actual and estimated values. Mean Squared Error (MSE) is the square of MAE, and Root Mean Squared Error (RMSE) is the square root of MAE. Root Mean Squared Log Error (RMSLE) is the logarithm of RMSE. R2, also known as the Coefficient of Determination or Goodness of Fit, evaluates the regression line's quality in comparison to the mean line, irrespective of context. The Pearson Correlation Coefficient measures the strength of the relationship between variables.

*Ranking tasks:*In the ranking context of recommendation systems, commonly utilized metrics include Hit Ratio (HR), Average Precision (AP), Mean Reciprocal Rank (MRR), Normalized Discounted Cumulative Gain (nDCG), and Discounted Cumulative Gain (DCG). Mean Reciprocal Rank (MRR) or Average Reciprocal Hit Ratio (ARHR) is derived by weighting the reciprocal rank of relevant scores among the top n entities and then summing them up. DCG represents the level of information gain within a prediction set by sorting documents based on relevance and position. However, since performance can't be compared based on a single query or criterion, nDCG scales the results considering the best among those analyzed.

*Clustering tasks:*When assessing clustering methods and determining the ideal number of clusters, several techniques are employed. These include the Average Silhouette Criterion, which measures an entity's similarity to its cluster in comparison to others. Calinski Harabasz [84] evaluates the sum of inert clusters and intra-cluster dispersions. The Elbow method charts the observed variation against the number of clusters (k) to identify the most suitable value. The V-score assesses the significance of the current clusters. Initially, [86] Homogeneity (h) illustrates the extent of similarity among samples within clusters, using Shannon's entropy. Following this, completeness (c) measures the accuracy of clustering similar samples. This involves calculating H(X/L) in the same way as H(L/X). To evaluate the algorithm's efficacy, Normalized Mutual Information (NMI) is computed between homogeneity and completeness for the V-measure.

*Group evaluation:*In cases necessitating group scores, several measures are utilized, such as Group Prediction Score [25], Group Divergence, Consensus Function, and T-Testing. The Group Prediction Score amalgamates individual-level predictions within a specific group. Group Divergence measures the dissimilarity in predictions at individual levels within a group. P-values gauge the likelihood of outcomes when the null hypothesis holds true. While they lack individual significance or a numeric score, they provide insight into the alignment of a score with a particular sample. T-testing [87] is employed to identify notable differences between the means of groups.

*Fairness:*Fairness in recommendations pertains to unbiased suggestions concerning individuals, groups, creators, or providers. Two primary types—Process and Outcome—are defined [75], with Outcome Fairness further subdivided into eight subclasses: individual, group, consistent, calibrated, counterfactual, envy-free, Rawlsian maximum, and maximum shared. Metrics used to evaluate these fairness types include Kolmogorov-Smirnov statistic, rKL, rRD, rND, Pairwise Ranking Accuracy Gap (PRAG), variance, Min-Max Difference (MMD), F-statistic (ANOVA), Gini coefficient, Jain's index, Min-Max Ratio, entropy, Least Misery, KL-divergence, MinSkew, MaxSkew, JS-divergence, Overall Disparity, L1 norm, and envy-freeness. Additionally, measurements like Equity Attention for group fairness (EAGF) and Supplier Popularity Deviation (SPD) are employed in this context.

*Subjective / Objective tasks:*Subjective measures focus on recognizing and evaluating user emotions consciously, while objective measures encompass both conscious and unconscious emotions. Objective measures [29] typically involve unconscious responses such as preferences for new or unexplored items, while subjective measures revolve around conscious emotions like user trust [29], self-reported feedback, and timely responses to questions.Objective measures include accuracy metrics, various estimations, user exposure to items, demographic relationships, and novelty relevant to users. Subjective measures, on the other hand, involve human judgments based on user ratings or questionnaires, as well as creators' perspectives like clicks, dwell time, shares, and dislikes.Metrics like Intra List Similarity (ILS), user coverage (UC), diversity, and novelty [83] are utilized. ILS calculates the average similarity among pairs based on specific metadata and context. Diversity is derived by subtracting user set similarity from one. Novelty involves considering overall rankings and averaging the popularity rankings of each user's top predictions.

*Offline / Online tasks:*For online evaluations, specific parameters like Click Through Rate (CTR), indicating the ratio of users clicking a link to those viewing it, reveal the online scenario. Dwell time measures how long a user spends on a link, while the bounce rate signifies the rate at which users return to the main engine or close a particular link. Additionally, metrics like average time spent on a link, session durations, and repetition in sessions contribute to online evaluation. Offline evaluations focus on measures such as DCG, nDCG, MRR, and MAP. These metrics are commonly used to assess the top k entities recommended from the available set based on user behavior history. However, they might not provide insights into new or alternative recommendations that a user would potentially like.

6 Results:

6.1 RQ1 –Investigation of Sparsity and long tail:

Despite extensive research spanning over a decade on recommender systems, the main challenge persists: sparsity significantly hampers these systems' performance. Sparsity occurs when there's an inadequate amount of meaningful transactions, making it challenging to accurately infer user similarities or identify relevant items. When users rate items with minimal feedback, finding similar users becomes challenging, impacting the accuracy of recommendations. Bridging this gap requires overcoming the obstacle of calculating precise predictions.

In recommender systems, the "long tail" phenomenon denotes recommending items with minimal interactions (ratings/likes). The challenge arises from a shortage of these less-popular items in recommendation lists. Over time, sales decline for many items, intensifying in the long term. Consequently, the system tends to duplicate popular items, resulting in a less engaging user experience. To address this, one approach involves segmenting the entire item collection into the "head" (popular) and "tail" (less popular) sections, focusing on clustering only the less-popular tail items.

Outlining above issue through an example, considering user_interactions and shared articles datasets, the initial step involved one-hot encoding the data to convert it into the required numerical format. A heatmap was generated to analyze correlations among variables. The analysis revealed a weak correlation between 'contented' and 'authorsessionid,' prompting their exclusion from further analysis. Subsequently, the datasets were analyzed to understand data sparsity and density, aiming to address issues like the new item problem and the long-tail problem. When items gain popularity, recommending them becomes simpler. However, discovering niche items to surprise users poses challenges due to limited historical data or specific interests provided. Hence, analyzing sparsity patterns in the data becomes crucial. The analysis displayed the count of points within each region. However, due to the vast dataset and multiple data points partially overlapping or sharing the same location on the graph, the issue of overplotting occurred, leading to an unclear visualization (Figure-6).

Fig 6 Heat map of feature-correlations

To gauge the correlation strength among numerical and categorical variables and identify optimal features for machine learning model construction, transforming the correlation matrix into color-coded visualizations is advantageous. The heatmap above showcases the behavior of various variables, including contentid, bookmark, comment created, follow, like, view, virality, timestamp, authorpersonId, authorsessionId, textWordCount, eventType_CONTENT REMOVED, eventType_CONTENT SHARED, eventType_HTML, contentType_RICH, and content_Type_video. This visualization aids in understanding the interrelationship and influence of these variables on each other, enabling informed decisions for model building.

Possible solutions include changing the transparency/opacity of the points by setting the alpha argument in geom_point() which can change the alpha argument to be any value between 0 and 1, where 0 sets the points to be 100% transparent and 1 sets the points to be 100% opaque. By default, alpha is set to 1. Next include jittering all points, i.e., giving each point small "nudge" in a random direction. Further if variables are categorical, as given FacetGrid() function of seaborn can be applied taking every dataframe as input, naming variable which in turn can form the row, column, or hue dimensions of the grid. One another solution include creation of multitude of small fragments and then representing number of points in this fragment. [112] Proposed gather plots which are defined as form of unit visualizations. This was successful in aggregation avoidance while maintaining the identity of individual entites. Still data sparsity and long tail problem has many sides to explore increasing its research scope. Figure- 7 shows sparsity in 8 datasets.

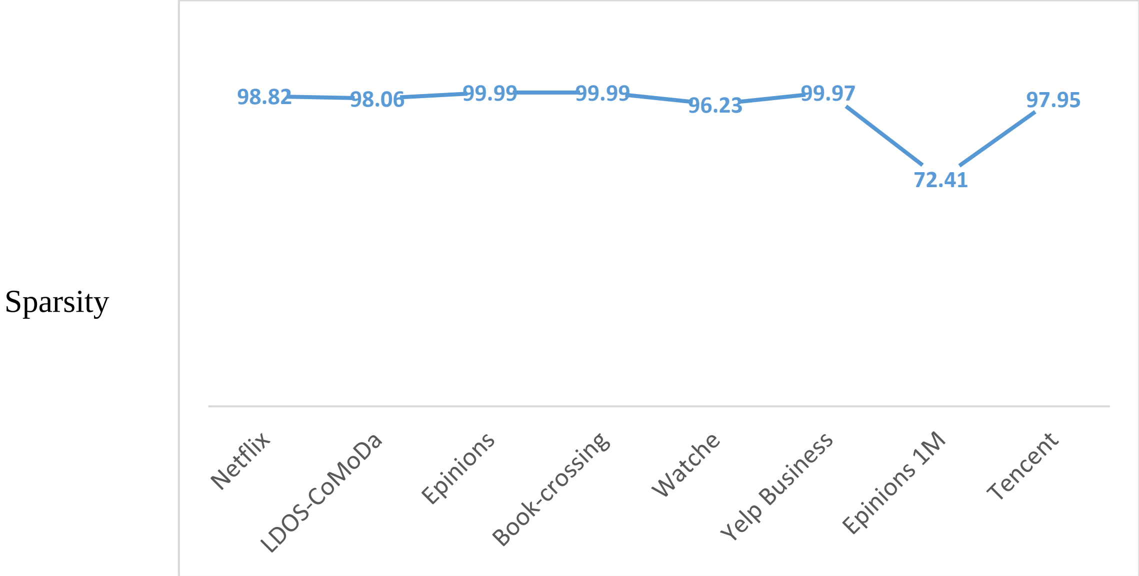

Fig-7 Sparsity of 8 commonly utilized datasets

6.2 RQ2 –Investigation of Multimodal data:

Multimodal recommendation systems aim to provide personalized suggestions by incorporating various user and item information, considering styles and aesthetic preferences. These systems utilize multiple forms of data to suggest items based on user input, history, and even visual aspects like color and pattern matching. However, integrating item multimodal content into existing sequential recommendation frameworks faces two key challenges [113]. Firstly, different types of content modalities often have distinct representations. Secondly, there's a disconnect between users' sequential behaviors and the multimodal content of items. Researchers have proposed solutions like transfer learning, topic modeling, and LDA to tackle these issues. These systems aim to recommend items aligned with a user's preferences, saving time and potentially increasing revenue by suggesting items with similar themes, styles, or ambiances.

Most recommender systems rely on a single type of data modality, often centered around text-based information such as feedback, ratings, and purchase history. These systems analyze and utilize textual data to generate recommendations personalized to user preferences and past interactions. But when it comes to document recommendations though excluding video, audio data, images, location and other social parameters too contribute to context. Demography and temporal features contribute to decision making focusing language, time and culture. Physiology plays an important role in mood and social preferences. The following Figure-8 shows consideration multimodal features in literature.

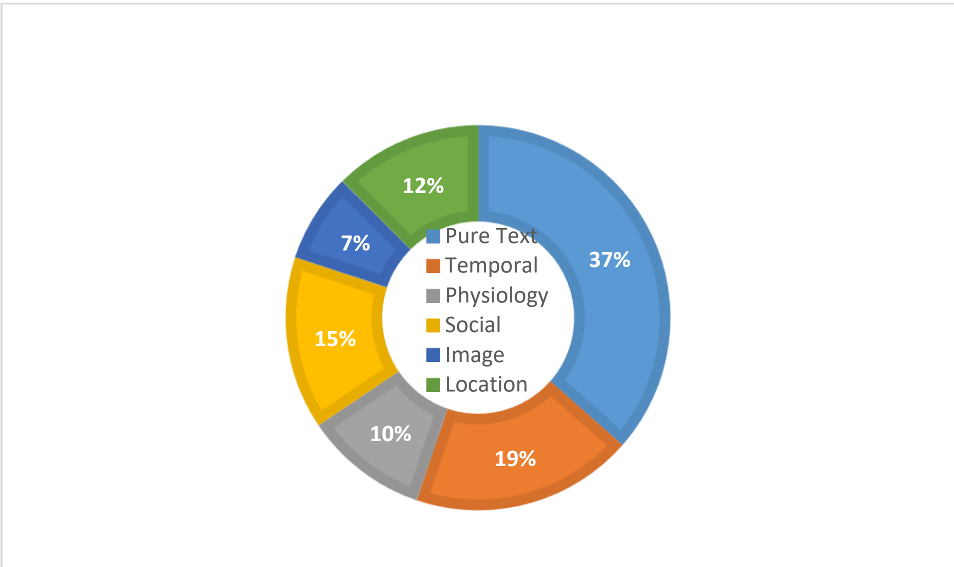

Fig-8 Percentage of multimodal features considered for textual recommendations.

6.3 RQ3 –Investigation of Cold start:

The cold start problem emerges when a recommender system lacks adequate data to make meaningful recommendations. This occurs when new users have rated very few or no items, posing a challenge in generating effective suggestions. Solutions to this issue include: (a) Prompting new users to specify their preferences explicitly. (b) Encouraging new users to rate items initially. (c) Gathering demographic data or metadata from users to tailor recommendations based on this information. The problem of weak trust, long tail or short head co exists along with cold start or warm start. This relation is depicted through Figure- 9.

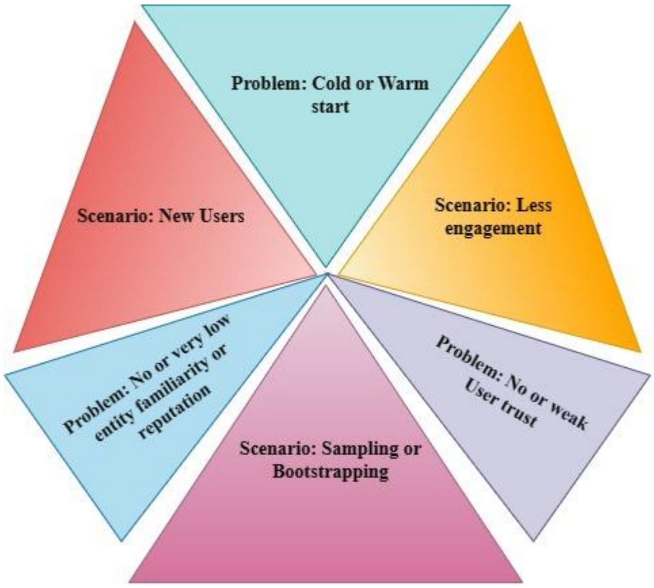

Fig- 9 Co-existence, relation and impact of weak trust, long tail and cols start.

By the fact that evaluation any recommender system requires measurement of ranking quality, the following Figure- 10 shows average nDCG performance concerning cold start users for 8 commonly used datasets.

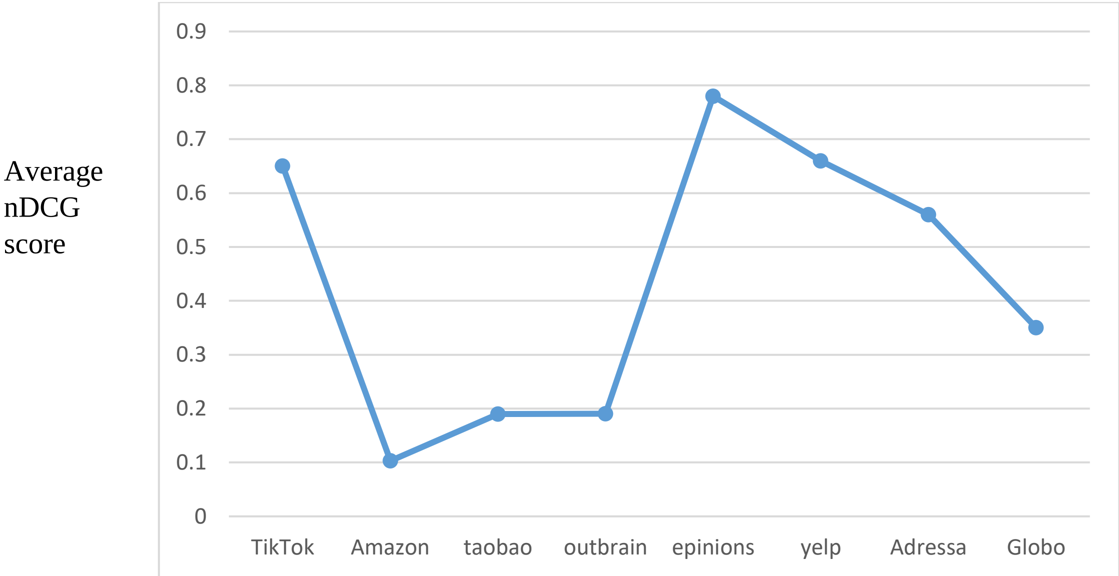

Fig- 10 Average nDCG performance concerning cold start users for 8 commonly used datasets

6.4 RQ4 –Investigation of Explainability within recommendation systems focuses on clarifying why an AI system made a specific decision, recommendation, or forecast. It aims to offer insights into the rationale behind the system's choices, benefiting users and system developers by elucidating why certain items are suggested. Achieving this involves comprehending the inner workings of the AI model and the data it relies on for training.It has two types- post-hoc and model-based. Post-hoc explainability provides clarifications after the model has made its decision, while model-based explainability integrates explanations within the model itself.In the realm of explainable recommendation systems, the goal is to create models that not only produce top-notch recommendations but also provide easily understandable explanations for those recommendations. Following Figure-11 shows level of goals achieved for explainability in recent literatures. And Figure-12 depicts how various interpretability factor lead to explainability.

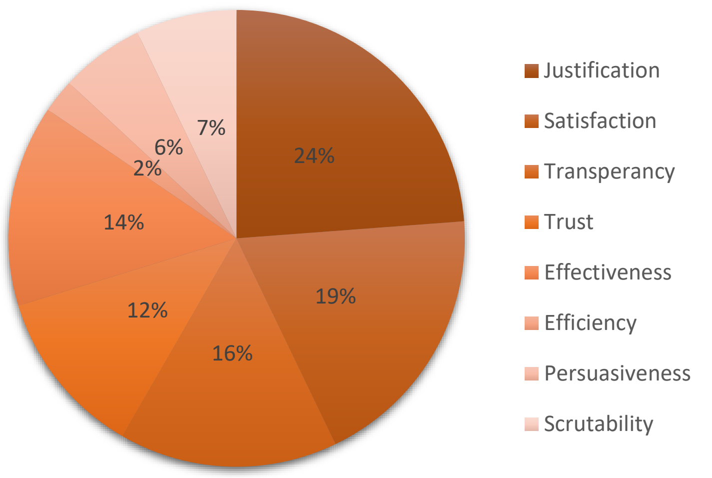

Fig- 11 shows level of goals achieved for explainability in recent literatures [114]

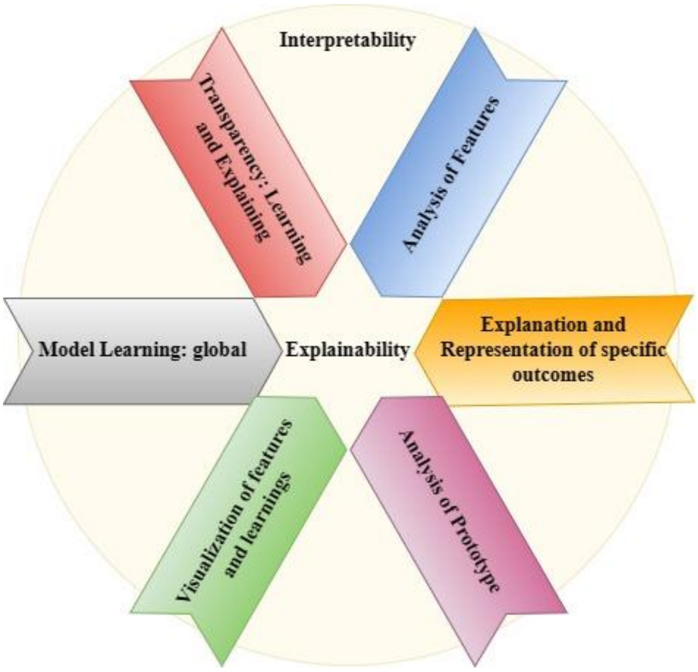

Fig-12 Various factors leading to explainability

6.5 RQ5: Investigation of Evaluation metrics utility

Assumingthe defined categories of evaluation metrics in section 5.2 namely Business, Online/ Offline, Objective/ Subjective, Fairness, Classification, Clustering, Regression, Ranking, Similarity distance and Group evaluations, Following Table shows the condensed overview of an extensive examination of evaluation metrics, each tailored to a specific task, highlights crucial factors to consider when utilizing these metrics. This presents a concise summary outlining the defining features and considerations essential for employing these evaluation metrics effectively in their respective tasks (Table-1).

Evaluation metric	Important aspects/ ideas to be taken care of	Task in recommendation
p-value	As more data is collected, less patterns can be inferred from it	Group evaluations
Kolmogorov-Smirnov (KS) statistic	Only for continuous distribution and more sensitive at center of distribution, measure high order inconsistencies	Fairness

EAGF	Increase in value ensures balanced recommendations	Fairness
SPD	Low value implies high performance	Calibrated Fairness
MRR	Handled according to implicit/ explicit data	Ranking
HR and recall	Proper selection of top k recommending list	Ranking
PCC	For quantitative variables, linear relations and normal distribution, cannot handle outlier, cant differentiate dependent and independent variables, not for homogeneous data	Recommendation as regression task
Accuracy	For a balanced problem where every class has equal importance.	Recommendation as classification task
F1	For binary classification and positive classes are more important	Recommendation as classification task
AUC curve	Gives equal importance to positive and negative classes, not for imbalanced data	Recommendation as classification task, ranking prediction
DCG	Can compare system (same parameters and recommending same number of items), checks whether independent variable contributes to model or not.	Ranking
nDCG	Unboundedness with negative samples i.e., do not penalize irrelevant estimations	Ranking
R^2	Cannot recognize bias in predictions	Recommendation as regression task
MAE	Ever error either small or large has equal weightage, cannot handle outlier	Recommendation as regression task
Average precision	For imbalanced data weighing positive samples more	Ranking
MSE	For normalized data, fluctuates in presence of outliers	Recommendation as regression task
RMSE	Sensitive to outliers	Recommendation as regression task
RMSLE	Can handle outlier, long tail exists for target value	Recommendation as regression task
MCC	Best value and worst value of 1 and -1 occurs for positive and negative correlation respectively.	Recommendation as classification task
CH index	Used when ground truth is unknown, high score means better performance, how good inherent features of data are used for creation of well clusters	Recommendation as clustering task
V score/ NMI	Measures whether independent label assigning methods of two datasets agrees or not in case where ground truth value is unknown	Recommendation as clustering task
t-test	For independent and normally distributed data, homogeneous variance	Group evaluations
ANOVA	Comparing mean of 2 or more groups/ comparison of multiple pairs	Group evaluations

Table-1 Summary of evaluation metrics

7. Challenges:

Privacy and Security: The data used in recommendation systems are vulnerable to attacks like online adversarial behavior and data poisoning. These attacks can insert deceptive or remove vital information, leading to skewed model parameters and biased training, influencing user exposure and market manipulation. Solutions proposed in studies [31] and [89] involve storing public data on a central server and keeping user data locally. They exchange necessary data to compute the model locally, securing the weights used in data exchange through Differential privacy. This approach limits information leakage but introduces communication overhead and remains susceptible to model attacks. The profiling phase presents a high-risk period for user information exposure, despite regulations like the General Data Protection Regulation (GDPR) not fully addressing misuse by public organizations, necessitating additional privacy regulations and guidelines.

Data Anonymisation: Anonymizing data is an effective solution for privacy concerns, yet it introduces challenges affecting accuracy. Initiatives like accessing cookies [91] to achieve this are hindered when users switch to incognito mode, complicating implementation. [93] Explored using anonymized data with homomorphic encryptions to enhance privacy and reduce third-party reliance. However, certain anonymization techniques struggle with high-dimensional data [92], requiring a delicate balance in considering attribute sensitivity. Employing anonymization must coincide with proper data masking and sensitive attribute categorization to maintain system performance [91]. Moreover, it's challenging to extract data patterns from anonymized data, crucial for recommender systems' insights. Additionally, while anonymization doesn't allow reidentification of encrypted entities, pseudonymization facilitates such identifiability.

Filter Bubbles and Bias: Customer behavior and consumption data serve as pivotal aspects in business analytics [85]. Recommendations often rely on past behavior and deductions drawn from it. However, this tendency can result in observation bias, fostering filter bubbles. These bubbles confine users to limited content that aligns with and reinforces their existing mindset and beliefs. Consequently, this can amplify self-reinforcing extremism and create echo chambers that dismiss alternative perspectives. Moreover, imbalanced data can introduce biases, such as commonly suggesting cricket, kabaddi, or sports-related content to males while recommending festival discounts or sales to females. While analyzing consumption behavior is essential, over-specialization in this regard can contribute to filter bubbles.

*Fairness:*The analysis generally indicates a conflict between system performance and accuracy when considering fairness [75]. However, achieving optimal solutions necessitates addressing all these aspects. While numerous works, like [90], delve into fairness, they often focus on a single type. Exploring joint, all-encompassing, and comprehensive fairness, with a detailed examination of the effects of different evaluation metrics, remains an unexplored area. To address fairness effectively, identifying sensitive attributes and their proper utilization is equally critical. Many existing studies assume that the necessary fairness-related attributes are present in the data, but there might be instances where this essential information is missing, underscoring the need for further investigation into fairness-related aspects. Implementing explainability in the recommendation task for fairness could persuade users more effectively, potentially enhancing their satisfaction levels and bolstering trust in the system [75]. For instance, in [32], adversarial learning was employed to discern bias-free (learning independent attributes of bias) and aware (sensitive attributes) embeddings of users, effectively reducing unfairness caused by bias. However, fine-tuning hyperparameters to eliminate bias and enhance fairness can be challenging and might expose the model to potential vulnerabilities. Additionally, it's important to note that fairness doesn't imply equal content exposure to all users; frequently active shoppers might receive more accurate estimations compared to those who rarely engage in shopping activities.

Serendipity: This concept combines various traits: relevance of an entity to a user (its connectedness or usefulness to the user), unexpected or surprising entities (those that seem fascinating or attention-grabbing), and novelty (items that the user hasn't encountered or been exposed to yet but may find interesting). For instance, if a user encounters technical content that they haven't seen before and reads it, it's considered novel. However, it doesn't qualify as serendipitous unless the user desires to engage with it further. Serendipity occurs when the user becomes interested in exploring that topic more after the initial encounter [29].

Time Factor: The demand for recommender systems involves responsiveness to user behavior within specific time frames. This necessitates real-time processing, low latency, efficient, and swift handling of large data volumes. Additionally, recommendations need to be timely, considering trends, popularity, seasonal patterns, individual preferences, recent inputs, and unique predictions. For instance, news articles tend to become outdated more quickly than technical content. Moreover, the duration of consumption or time spent on each varies significantly—news articles typically receive less attention compared to the prolonged engagement seen with technical content.

Duplications: This issue arises when numerous links present similar content, causing users to lose interest or trust in the system. Conversely, some links may actually offer valuable new information, but users might mistakenly perceive them as irrelevant or untrustworthy. As a result, users may be confined to a limited range of content, reducing their exposure to diverse information. To address this problem, providing explanations for recommendations can be an effective solution.

Diversity: To enhance user experience, trust, and engagement, providing a diverse array of information is crucial. When users encounter similar links or content, they tend to lose interest. However, many studies primarily focus on optimizing recommendation accuracy while neglecting the importance of diverse content [43]. Diverse recommendations not only involve suggesting new content different from previous recommendations but also expose users to a varied range of top n interests. Diversity measures the uniqueness and variability among predicted items, often implemented through re-ranking techniques like Intra list similarity, normalized diversity, or temporal diversity [29]. However, employing diversity through re-ranking poses a trade-off between accuracy and diversity due to high computational demands. Future directions in diversity research involve exploring the accuracy-diversity trade-off, assessing diversity levels tailored to individual users [29], and designing scalable systems for diversification.

Incognito tabs: Handling user data in incognito tabs poses a challenge for recommendation systems as it restricts access to essential user information stored in cookies, history logs, and tracking. Due to the limited updates in users' interests, it may lead to repetitive recommendations, potentially causing user disengagement. Future efforts should focus on devising methods to identify and manage user data when incognito tabs are utilized before employing recommendation systems for predictions.

*Content Validation and Control:*The prevalence of spam or fraudulent content can mislead users, necessitating that all recommended data be current, authenticated, and sourced from authorized channels. It's crucial to filter out low-quality or harmful content before making recommendations, addressing the challenge of misinformation or content that can have adverse social impacts.

Explainability: It is equally necessary to provide user with the reason of any entity being recommended to them in order to boost and maintain their trust for system. The deep learning work flows and outputs are really not easily understandable by non technical users. [47] Analyses the use of attention based and graph based models to provide explainability, as weights of attention modules gives intuition about internal working which in turn can explain feature contribution for model. Providing information about interactions, entity relations, semantics, rules and associations, knowledge graph can solve issues with cold start, explainability and accuracy. Integration of visualization for recommendation tasks improves explainability [88].To the best of knowledge, a system that can give user with proper overall explanations of recommendations hasn't been developed yet. However, providing proper explanations in an online environment still remains unexplored and has a vital scope for future work.

*Long Tail:*The term refers to the situation where only short head items (set of purchased and popular entities) are recommended but long tail (set of purchased but unpopular and less viewed) aren't introduced to users [88]. [94] Categorized this problem to usage of clusters, graphs, deep learning, ranks, linear model, relevance, user –values, multi-level item similarity and multiple evaluation metrics. But while considering users to be within session or sequencing scenarios, problem exists for balancing short heads-long tails and recommendation decision made for them. Second issue with these, is consideration of only entity ratings, user and entity features, other sensitive information like demographics, time factor and semantics or context are being ignored. Finally for evaluation purpose metrics like diversity, coverage rate (percentage of discrete items among recommendations) and popularity distributions are considered commonly excluding serendipity. Long tail recommendation need exploration of proper evaluation metrics according to data, environment and users contexts.

*Multimodal and Multilingual data:*The domain of document recommendation encompasses a variety of content formats, including images, links, and texts, often spanning multiple languages [29]. However, current research predominantly focuses on single data types—text, images, or videos—failing to address the challenges arising from the fusion of multiple formats. For instance, technical document recommendations may involve not only textual information but also images (like workflows or equations) and links (directing to codes or videos), necessitating comprehensive approaches that encompass diverse content types.

Other Issues: Sequential consumption is beneficial for users seeking updated or related information in subsequent readings. However, this approach might not be favorable when it repetitively suggests similar items, as seen in music or movie recommendations [29]. Addressing the lack of a benchmark dataset capable of measuring comprehensive metrics, including "total fairness" for all system stakeholders, remains a challenge. Additionally, understanding user churn rates offers insights into satisfaction levels, the speed of recommendation changes post new ratings, and underlying interests. Lowering churn rates can foster user trust. Besides accuracy metrics, effectively managing missing data and noise is crucial. A system's ability to provide relevant suggestions despite noise and outliers showcases its robustness. Moreover, addressing challenges related to data sparsity and cold start problems when new entities, whether users or items, join the online environment requires further exploration.

8. Discussions and Future Scope:

Deep Reinforcement Learning: The implementation of deep reinforcement learning offers valuable insights into model operations, shedding light on consumer behaviors and decision-making rationales. This visibility aids in monitoring the training phase, enabling prompt intervention if any undesirable occurrences arise. However, such methods allow a deep scrutiny of internal states, rendering them unsuitable in scenarios prioritizing security and privacy concerns. They find optimal application in unstructured environments. Leveraging SHAP (SHapley Additive exPlanations) values allows for an analysis of long-term effects concerning the combination of state and action for each user. RL-SHAP plotting effectively illustrates the impact of features on actions, delineating whether these effects are positive, negative, high, or low.

Interpretability and Explainability: When security takes precedence, machine learning (ML) or deep learning (DL) models are often non-transparent. This lack of transparency impedes comprehensive understanding, making it challenging for developers to identify failures and their root causes. To address this, various frameworks such as SHAP, LIME (Local Interpretable Model-agnostic Explanations), Shapash, InterpretML, ELI5, and OmniXAI (Omni eXplainable AI) have emerged to enhance explainability and interpretability in these models.

Blockchain: There's a distinct gap between model performance and its security. Blockchain technology, known for its attributes such as resiliency, durability, flexibility, and trust, offers potential contributions to privacy-preserving recommender systems by eliminating third-party interference. It's seen as a solution to common challenges like transparency and decentralized training, ensuring security. However, challenges like scalability, distributed training, and explainability need careful consideration when employing blockchain in these systems [96].

Multitask and Transfer Learning: Recommendation tasks often involve multiple objectives, necessitating the optimization of multiple losses. Multi-task learning (MLT) facilitates this optimization by allowing variables to be shared among tasks and enabling transfer learning. Previous work has highlighted the use of transfer learning (TL) in addressing data sparsity issues [29]. However, addressing noise, outliers, and missing data requires considering strategies to handle the differences in dataset features.

Federated Learning: Decentralized management becomes necessary due to hardware limitations for online computations. On-device computations may be constrained by data, labeling, and model size limitations. Federated Learning (FL) has effectively addressed data sensitivity concerns in handling heterogeneous devices, edge computations, and their optimization. FL holds substantial potential in the online environment if challenges such as system heterogeneity, statistical variations, privacy concerns, and communication overhead are addressed efficiently [97].

Augmentation and Contrastive learning: Image features within document recommendations are often overlooked, and their adaptation to specific environments is rarely explored. Self-supervised learning, such as Contrastive Learning (CL), could prove effective in an online environment due to its label-free learning nature. However, the potential of CL in this particular recommendation task remains largely unexplored.

Raw Data Exploitation: Self-supervised methods have gained prominence for their ability to explore and analyze raw data directly, providing richer prompts for effective model training. It begins with a thorough examination of augmentation techniques for pre-training. Leveraging augmentation can facilitate learning without the reliance on labels. Once augmentation is addressed, the focus shifts to the primary task of making recommendations. However, leveraging the pre-trained model with augmentation can potentially enhance the performance of recommendation systems even further.

Multi-objective reinforcement learning (MORL): Multi-Objective Reinforcement Learning (MORL) is designed to address multiple goals simultaneously. While Reinforcement Learning (RL) is adept at handling uncertainties in sequential estimations, it typically maximizes selected numerical values representing a single long-term objective. Even when considering multiple objectives, they are often viewed as a linear combination. Factors to consider when conceptualizing solutions include the nature of single or multiple policies, the nature of utility functions, deterministic or stochastic policies, scalarized expected returns (SER), and expected scalarized returns (ESR) according to [99]. This powerful technique can make systems more adaptable in real-time scenarios.

Causal Bayesian networks (CBN): Metrics used for fairness often focus on the relationship between sensitive features and the output. However, training data patterns can generate unfairness that might affect other attributes as well. Causal graphs or CBNs (Causal Bayesian Networks) can aid in understanding how sensitive attributes impact others in the process. Constructing causal graphs poses a primary challenge, often relying on general information such as ID details [75]. Including more features complicates the construction process. Furthermore, addressing extraneous determinants and managing the complexities of conducting counterfactual inference [98] are challenging aspects. There's significant scope for further exploration and improvement in addressing these drawbacks.

9. **Conclusion:**

This study focuses on different classifications of document recommendation tasks and the features considered within each category. We delve into recent research and literature pertaining specifically to document recommender systems, analyzing and comparing techniques and models based on categorized tasks and their respective key features. Additionally, we present comparisons of the various goals targeted by literature in each specific category. The comprehensive overview includes insights into datasets and evaluation metrics. Through an analysis of these metrics, we identify key points to enhance the responsibility of recommendation systems. Results section provide insights to the quality of this study. By uncovering challenges and suggesting unrealized potential, we provide the latest perspectives and insights for future research in this field. The knowledge shared in this work aims to inspire researchers to develop progressive solutions for the current domain and other real-time application domains.

List of Abbreviations:

CTR: Click Through Rate

P-LDRS: Pre-Learned Legal Document Recommender System

ACM: Association for Computing Machinery

GRU: Gated Recurrent Unit

GDELT International Press Telecommunications Council

IPTC: International Press Telecommunications Council

MIND: Microsoft news dataset

SMOTE: Synthetic Minority Oversampling Technique

RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers

KDM: Keyword Diversity Measurement

LDA: Latent Dirichlet Allocation

PMF: Probabilistic matrix factorization

SCA: Sentence Complexity Analysis

SQM: Scientific Quality Measurement

CBF content-based filtering

RARD: Related-Article Recommendation Dataset

MCC: Matthews Correlation Coefficient

MRR: Mean Reciprocal Rank

rKL: Normalized discounted KL-divergence

ILS: Intra List Similarity

UC: User Coverage

SPD: Supplier Popularity Deviation

MMD: Min-Max Difference

PRAG: Pairwise Ranking Accuracy Gap

ANOVA: Analysis of Variance

nDCG: normalized Discounted Cummulative Gain

RMSE: Root mean squared error

MAE: Mean absolute error

MSE: Mean squared error

BS: Brier score

BPR: Bayesian personalized ranking

GLM: Generalized linear modeling

CAOT: Citation Analysis Over Time

SEAN: Social Explorative Attention Network

ORP : Open Recommendation Platform

CF: collaborative filtering

MF: matrix factorization

NMF: Non-negative matrix factorization

lg2vec: log2vec

AUC: Area Under Curve

ROC: Receiver operating characteristic

PCC: Pearson Correlation Coefficient

HR: Hit ratio

AP: Average precision

DCG: Discounted Cummulative Gain

ARHR: Average Reciprocal Hit Ratio

IDCG: Ideal Discounted Cummulative Gain

CH: Calinski Harabasz index

rND: Normalized discounted difference

EAGF: Equity Attention for group fairness

MORL: Multi-objective reinforcement learning

CBN: Causal Bayesian networks

References:

[1] Jenish Dhanani, Rupa Mehta, Dipti Rana. 2022. Effective and scalable legal judgment recommendation using pre-learned word embedding. Complex & Intelligent Systems, Springer. https://doi.org/10.1007/s40747-022-00673-1.

[2] Dhanani, J., Mehta, R. and Rana, D.P. 2021. Legal document recommendation system: a dictionary based approach. International Journal of Web Information Systems, Vol. 17 No. 3, pp. 187-203. https://doi.org/10.1108/IJWIS-02-2021-0015

[3] Nithirun Numnonda, Sapa Chanyachatchawan & Nuengwong Tuaycharoen, 2022. Journal Recommendation System for Author Using Thai and English Information from Manuscript. Lecture Notes in Networks and Systems book series (LNNS,volume 453), International Conference on Computing and Information Technology IC2IT 2022: Proceedings of the 18th International Conference on Computing and Information Technology (IC2IT 2022), pp 142–151, 03 April 2022

[4] Rui Wang; Yuliang Shi. 2022 Research on application of article recommendation algorithm based on Word2Vec and Tfidf. IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA).

[5] Stefano Ferilli , and Liza Loop. 2022. Toward Reasoning-based Recommendation of Library Items – A Case Study on the e-Learning Domain. 18th Italian Research Conference on Digital Libraries (IRCDL 2022), Padua, Italy, CEUR Workshop Proceedings.

[6] Choi Jaewoong, Lee Jiho, Yoon Janghyeok, Jang Sion, Jaeyoung Kim & Sungchul Choi. 2022. A two-stage deep learning-based system for patent citation recommendation. Scientometrics, Springer.

[7] Huang ZhengWei, Min JinTao, Yang YanNi, Huang Jin & Tian Ye. 2022 Recommendation method for academic journal submission based on doc2vec and XGBoost. Scientometrics,
Springer. Volume 127, pages2381–2394.

[8] Duc H. Le, Tram T. Doan, Son T. Huynh, and Binh T. Nguyen. 2022 SimCPSR: Simple Contrastive Learning for Paper Submission Recommendation System. arXiv:2205.05940v1 [cs.IR].

[9] Aram Karimi, Simon Dobnik. 2022. Multi-task recommendation system for scientific papers with high-way networks. arXiv:2204.09930v1 [cs.IR].

[10] Nianlong Gu, Yingqiang Gao & Richard H. R. Hahnloser. 2022. Local Citation Recommendation with Hierarchical-Attention Text Encoder and SciBERT-Based Reranking. Lecture Notes in Computer Science book series (LNCS,volume 13185). European Conference on Information Retrieval. ECIR 2022: Advances in Information Retrieval pp 274–288.

[11] Gauri Chaudharya, and Manali Kshirsagar. 2021 Enhanced Text Clustering Approach using Hierarchical Agglomerative Clustering with Principal Components Analysis to Design Document Recommendation System. Part of Research Transcripts in Computer, Electrical and Electronics Engineering. Vol: 02, Advanced Research in Computer Engineering. Grinre.

[12] Vilmos Tibor Salamon, Tsegaye Misikir Tashu & Tomáš Horváth. 2021. Linear Concept Approximation for Multilingual Document Recommendation. Part of the Lecture Notes in Computer Science book series (LNISA,volume 13113). Intelligent Data Engineering and Automated Learning – IDEAL 2021 pp 147–156.

[13] Panyadee, P., Wita, R. 2021. Integrating Graph-Based Document Recommendation in Digital Libraries of Theses Collection. In: Meesad, P., Sodsee, D.S., Jitsakul, W., Tangwannawit, S. (eds) Recent Advances in Information and Communication Technology 2021. IC2IT 2021. Lecture Notes in Networks and Systems, vol 251. Springer, Cham. https://doi.org/10.1007/978-3-030-79757-7_22.

[14] Ansel MacLaughlin , Tao Chen , Burcu Karagol Ayan, Dan Roth. 2021. Context-Based Quotation Recommendation. Association for the Advancement of Artificial Intelligence. Google Research.

[15] Wenyi Xiao, Huan Zhao, Haojie Pan, Yangqiu Song, Vincent W. Zheng, Qiang Yang. 2020. Social Explorative Attention based Recommendation for Content Distribution Platforms. Data Mining and Knowledge Discovery (https://doi.org/10.1007/s10618-020-00729-1). arXiv:2012.04945v1 [cs.SI] 9 Dec 2020.

[16] Papneja, S., Sharma, K., Khilwani, N. 2020. Content Recommendation Based on Topic Modeling." In: Singh, V., Asari, V., Kumar, S., Patel, R. (eds) Computational Methods and Data Engineering. Advances in Intelligent Systems and Computing, vol 1227. Springer, Singapore. https://doi.org/10.1007/978-981-15-6876-3_1.

[17] Akshay Gugnani, Hemant Misra. 2020. Implicit Skills Extraction Using Document Embedding and Its Use in Job Recommendation. Implicit Skills Extraction Using Document Embedding and Its Use in Job Recommendation, IBM Research.

[18] Rrezart Prebreza, Daniela Gotseva, Ognyan Nakov. 2021. Recommendation Systems Based on Textual Document Analysis. 29th National

Conference with International Participation (TELECOM), IEEE.

[19] Donghoon Lee, Namgyu Kim. 2021. Methodology for Search Intent-based Document Recommendation. Journal of The Korea Society of Computer and Information Vol. 26 No. 6, pp. 115-127.

[20] S. Renuka, G. S. S. Raj Kiran & Palakodeti Rohit. 2021. An Unsupervised Content-Based Article Recommendation System Using Natural Language Processing. Data Intelligence and Cognitive Informatics pp 165–180, Springer.

[21] Khan Omer. 2021. Knowledge Management: Document Similarity Based Recommendation. Aalto University.

[22] Mr. Dhiraj Vaibhav Bagul, Dr. Sunita Barve. A novel content-based recommendation approach based on LDA topic modeling for literature recommendation. Proceedings of the Sixth International Conference on Inventive Computation Technologies [ICICT 2021] IEEE Xplore Part Number: CFP21F70-ART; ISBN: 978-1-7281-8501-9.

[23] Sachin Papneja, Kapil Sharma, and Nitesh Khilwani. 2021. Content Recommendation Based on Topic Modeling. Computational Methods and Data Engineering, Advances in Intelligent Systems and Computing 1227, Springer Nature Singapore Pte Ltd.

[24] Raghvendra Rao, Suyash Choubey, Georg Gutjahr, and Prema Nedungadi. 2020. Document Recommendation for Medical Training Using Learning to Rank. Lecture Notes in Networks and Systems, Volume 141, Machine Learning for Predictive Analysis-. Proceedings of ICTIS 2020 Springer.

[25] Yuezhong Wu, Qiang Liu, Rongrong Chen, Changyun Li, and Ziran Peng. 2020. A Group Recommendation System of Network Document Resource Based on Knowledge Graph and LSTM in Edge Computing. Wiley, Hindawi, Security and Communication Networks Volume 2020.

[26] Ming Li , Ying Li , Wangqin Lou , Lisheng Chen. A hybrid recommendation system for Q&A documents. Expert Systems with Applications, 144(), Elsevier.

[27] Arpita Chaudhuri, Nilanjan Sinhababu, Monalisa Sarma, Debasis Samanta. 2021. Hidden features identification for designing an efficient research article recommendation system. International Journal on Digital Libraries. Part of Springer Nature.

[28] Yijian Hu, Shih-Yu Chang, Kai-Rui Hsu, Jaydeep Chakraborty, Srividya K. Bansal.2021 Ontology-based Document Recommendation System using Topic Modeling. IEEE 15th International Conference on Semantic Computing (ICSC).

[29] Shaina Raza & Chen Ding. 2022. News recommender system: a review of recent progress, challenges, and opportunities", Artif Intell Rev 55. Springer 749–800.

[30] Sungmin Cho, Hongjun Lim, Keunchan Park, Sungjoo Yoo, Eunhyeok Park. 2021. On the Overlooked Significance of Underutilized Contextual Features in Recent News recommendation Models. arXiv preprint arXiv:2112.14370.

[31] Chuhan Wu, Fangzhao Wu, Tao Qi, Yongfeng Huang. 2021. FeedRec: News Feed Recommendation with Various User Feedbacks. arXiv preprint arXiv:2102.04903.

[32] Chuhan Wu, Fangzhao Wu, Xiting Wang, Yongfeng Huang, Xing Xie. 2021. FairRec:Fairness-aware News Recommendation with Decomposed Adversarial Learning. In AAAI.

[33] Xuanyu Zhang, Qing Yang, Dongliang Xu. 2021. Combining Explicit Entity Graph with Implicit Text Information for News Recommendation. In Companion Proceedings of WWW.

[34] Peitian Zhang, Zhicheng Dou, Jing Yao. 2021. Learning to Select Historical News Articles for Interaction based Neural News Recommendation arXiv preprint arXiv:2110.06459.

[35] Songqiao Han, Hailiang Huang, Jiangwei Liu. 2021. Neural news recommendation with event extraction. arXiv preprint arXiv:2111.05068.

[36] Qi Zhang, Jingjie Li, Qinglin Jia, Chuyuan Wang, Jieming Zhu, Zhaowei Wang, Xiuqiang He. 2021. UNBERT: User-News Matching BERT for News Recommendation. In IJCAI. Zhi-Hua Zhou.

[37] Tao Qi, Fangzhao Wu, Chuhan Wu, Yongfeng Huang. 2021. Personalized News Recommendation with Knowledge- aware Interactive Matching. In SIGIR. Microsoft Research Lab.

[38] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2021. Is News Recommendation a Sequential Recommendation Task?. arXiv e-prints.

[39] Zhiming Mao, Xingshan Zeng, Kam-Fai Wong. 2021. Neural News Recommendation with Collaborative News Encoding and Structural User Encoding. In EMNLP.

[40] Mingyuan Ma, Sen Na, Hongyu Wang, Congzhou Chen, Jin Xu. 2021. The graph-based behavior-aware recom mendation for interactive news. Applied Intelligence (2021).

[41] Qinglin Jia, Jingjie Li, Qi Zhang, Xiuqiang He, Jieming Zhu, "RMBERT: News Recommendation via Recurrent Reasoning Memory Network over BERT. IN. SIGIR, Microsoft Research Lab.

[42] Dai Hoang Tran, Salma Hamad, Munazza Zaib, Abdulwahab Aljubairy, Quan Z Sheng, Wei Emma Zhang, Nguyen H Tran, Nguyen Lu Dang Khoa,. 2021. Deep News Recommendation with Contextual User Profiling and Multifaceted Article Representation. In WISE. Springer.

[43] Tao Qi, Fangzhao Wu, Chuhan Wu, Yongfeng Huang. 2021. PP-Rec: News Recommendation with Personalized User Interest and Time-aware News Popularity. In ACL, 5457-5467.

[44] Tao Qi, Fangzhao Wu, Chuhan Wu, Peiru Yang, Yang Yu, Xing Xie, Yongfeng Huang. 2021. HieRec: Hierarchical User Interest Modeling for Personalized News Recommendation. In ACL.

[45] Chuhan Wu, Fangzhao Wu, Tao Qi, Yongfeng Huang. 2021. MM-Rec: Multimodal News Recommendation. arXiv preprint arXiv:2104.07407.

[46] Jiasen Lu, Dhruv Batra, Devi Parikh, Stefan Lee. 2019. Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. Part of Advances in Neural Information Processing Systems 32 (NeurIPS).

[47] Andreea Iana, Mehwish Alam, Heiko Paulheim, 2022. A Survey on Knowledge-Aware News Recommender Systems. Semantic Web Journal.

[48] Alireza Gharahighehi, Celine Vens, Konstantinos Pliakos. 2021. Fair multi-stakeholder news recommender system with hypergraph ranking. Information Processing and Management 58 102663. Elsevier.

[49] Jon Atle Gulla, Lemei Zhang, Peng Liu, Özlem Özgöbek, and Xiaomeng Su. 2017. The Adressa dataset for news recommendation. In Proceedings of the International Conference on Web Intelligence (WI '17). Association for Computing Machinery, New York, NY, USA, 1042–1048. https://doi.org/10.1145/3106426.3109436.

[50] Benjamin Kille, Frank Hopfgartner, Torben Brodt, Tobias Heintz, 2013. The plista Dataset. NRS'13 Hong Kong, China. ACM 978-1-4503-2302-4/13/10.

[51] DanB , Rachael Tatman , Alexis Cook . 2019. Hacker News. Retrieved December 20, 2022 from https://www.kaggle.com/datasets/hacker-news/hacker-news

[52] Yahoo Labs. 2017. Language Data. Retrieved December 20, 2022 from https://webscope.sandbox.yahoo.com/catalog.php?datatype=l&guccounter=1

[53] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu and Ming Zhou. 2020. MIND: A Large-scale Dataset for News Recommendation. ACL.

[54] Wu et al. 2020. MIND: Microsoft News dataset. A Large-scale Dataset for News Recommendation. Retrieved December 20, 2022 from https://msnews.github.io/

[55] Sumanthvrao, Kaggle Kerneler. 2019. Fake-News-Dataset. Two fake news datasets covering seven different news domains. Retrieved December 20, 2022 from https://www.kaggle.com/datasets/sumanthvrao/fakenewsdataset.

[56] Bai, X., Wang, M., Lee, I., Yang, Z., Kong, X., Xia, F. 2019. Scientific paper recommendation: a survey. IEEE Access. 7, 9324–9339.

[57] Michael Färber & Adam Jatowt. 2020. Citation recommendation: approaches and datasets. International Journal on Digital Libraries volume 21, Springer. pages375–405.

[58] Palanigurupackiam Nagaraj, Perumalsamy Deepalakshmi. 2022. An intelligent fuzzy inference rule-based expert recommendation system for predictive diabetes diagnosis. Internatioanl Journal of Imaging systems and Technology, Wiley.

[59] Ghosh, Chinmoy and Saha, Subrata and Saha, Soumyadeep and Ghosh, Niladitya and Singha, Kumaresh and Banerjee, Annwesha and Majumder, Sourav. 2020. Machine Learning Based Supplementary Prediction System Using K Nearest Neighbour Algorithm. Proceedings of Industry Interactive Innovations in Science, Engineering & Technology (I3SET2K19). Available at SSRN: https://ssrn.com/abstract=3517197

[60] Isabel Hilliger, Tinne De Laet, Valeria Henríquez, Julio Guerra, Margarita Ortiz-Rojas, Miguel Ángel Zuñiga, Jorge Baier & Mar Pérez-Sanagustín. For Learners, with Learners: Identifying Indicators for an Academic Advising Dashboard for Students. European Conference on Technology Enhanced Learning, EC-TEL 2020: Addressing Global Challenges and Quality Education. pp 117–130, LNISA,volume 12315, Springer.

[61] Yi Li a , Ronghui Wang a , Guofang Nan a, Dahui Li b , Minqiang Li. A personalized paper recommendation method considering diverse user preferences. Decision Support Systems,Volume 146, Elsevier. 113546

[62] Satvik Garg. 2021. Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning. 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) IEEE.

[63] Kevin Appadoo, Muhammad Bilaal Soonnoo, Zahra Mungloo-Dilmohamud. 2020. JobFit: Job Recommendation using Machine Learning and Recommendation Engine. CSDE, IEEE.

[64] Zhengxiao Du, Jie Tang, Yuhui Ding. 2021. POLAR++: Active One-Shot Personalized Article Recommendation. IEEE Transactions on Knowledge and Data Engineering.

[65] Yanping Zhang et al. 2021. Citation Recommendation based on Citation links Explanation. J. Phys.: Conf. Ser. 1827 012069

[66] Danyang Liu, Jianxun Lian, Zheng Liu, Xiting Wang, Guangzhong Sun, Xing Xie. 2021. Reinforced Anchor Knowledge Graph Generation for News. KDD '21: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining

[67] IremIslek, -Sule GunduzOguducu. 2022. A hierarchical recommendation system for E-commerce using online user reviews. Electronic Commerce Research and Applications, Elsevier. Volume 52. 101131

[68] JIZHOU HUANG, HAIFENG WANG and WEI ZHANG, TING LIU. 2020. Multi-Task Learning for Entity Recommendation and Document Ranking in Web Search. ACM Transactions on Intelligent Systems and Technology. Vol. 11, No. 5, Article 54.

[69] Zhan Yang a ,Wei Xu a, Runyu Chen. 2021. A deep learning-based multi-turn conversation modeling for diagnostic Q&A document recommendation. Information Processing and Management, Elsevier.

[70] XinMei, XiaoyanCai, SenXu, WenjieLi, ShiruiPan, LibinYang. 2022. Mutually reinforced network embedding: An integrated approach to research paper recommendation. Expert Systems with Applications. Volume 204, Elsevier. 117616

[71] Viral Sheth & Kostadin Damevski. 2022. Grouping related stack overflow comments for software developer recommendation. Automated Software Engineering. volume 29. Article number: 40.

[72] Nicola Capuano, Luigi Lomasto, Andrea Pozzi & Daniele Toti. 2022. Natural Language Understanding for the Recommendation of Learning Resources Within Student Collaboration Tools. TLIC 2022: Innovative Approaches to Technology-Enhanced Learning for the Workplace and Higher Education. pp 483–495.

[73] Jenish Dhanani. Legal document recommendation system: A cluster based pairwise similarity computation. Soft Computing and Intelligent Systems: Techniques and Applications. Journal of Intelligent & Fuzzy Systems. vol. 41, no. 5, pp. 5497-5509.

[74] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Jundong Li, Zi Huang. 2022. Self-Supervised Learning for Recommender Systems: A Survey. IEEE Transactions on Knowledge And Data Engineering.

[75] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, Shaoping Ma. 2022. A Survey on the Fairness of Recommender Systems. ACM Trans. Inf. Syst.

[76] Christopher Brown. 2018. Refseer data. Retrieved December 20, 2022. from https://github.com/chbrown/refseer

[77] Cornell University, Shivanand Roy, et, al. 2023. arXiv Dataset. arXiv dataset and metadata of 2.2M+ scholarly papers across STEM. Retrieved February 6, 2023. from https://www.kaggle.com/datasets/Cornell-University/arxiv

[78] Lynx. JRC-Acquis - Dataset. Retrieved February 6, 2023. from http://data.lynx-project.eu/it/dataset/jrcacquis

[79] Wikipedia. AMiner (database). 2022. Retrieved February 6, 2023. from https://en.wikipedia.org/wiki/AMiner_(database).

[80] Paperswithcode. DBLP (Citation Network Dataset) Retrieved February 6, 2023. from

https://paperswithcode.com/dataset/dblp#:~:text=The%20DBLP%20is%20a%20citation,year%2C%20venue%2C%20and%20title.

[81] Papers With Code. Skill2vec. 2017. Retrieved February 6, 2023. from https://paperswithcode.com/dataset/skill2vec

[82] The Dataverse Project. RARD: The Related-Article Recommendation Dataset [Data] - Mr. DLib Dataverse. 2017. Retrieved February 6, 2023. From https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/HA8EAH

[83] Rishabh Bhatia, Medium. Recommendation System Evaluation Metrics. 2019. Retrieved February 6, 2023. from https://medium.com/@rishabhbhatia315/recommendation-system-evaluation-metrics-3f6739288870

[84] Misha Sv, Towards Data Science. Calinski-Harabasz Index for K-Means Clustering Evaluation using Python. 2022 Retrieved February 6, 2023. from https://towardsdatascience.com/calinski-harabasz-index-for-k-means-clustering-evaluation-using-python-4fefeeb2988e

[85] Raval, A. S., Oza, R. K., & Jain, A. V. Social Media Content Classification and Detection for Trend Analysis Using Machine Learning Techniques.

[86] Filippo Valle, Towards Data Science. V measure: an homogeneous and complete clustering. Retrieved February 6, 2023. from https://towardsdatascience.com/v-measure-an-homogeneous-and-complete-clustering-ab5b1823d0ad

[87] prakharr0y, sooda367, abhishek0719kadiyan, GeeksforGeeks. T-test. 2021. Retrieved February 6, 2023. From https://www.geeksforgeeks.org/t-test/

[88] Raval, A., Borisagar, K. 2022. A Survey on Techniques and Methods of Recommender System. In: Kalinathan, L., R., P., Kanmani, M., S., M. (eds) Computational Intelligence in Data Science. ICCIDS 2022. IFIP Advances in Information and Communication Technology, vol 654. Springer, Cham. https://doi.org/10.1007/978-3-031-16364-7_8

[89] Jia-YunJiang, Cheng-TeLi, Shou-DeLin. 2019. Towards a more reliable privacy-preserving recommender system. Information Sciences, Elsevier. Volume 482. Pages 248-265

[90] Yao Wu, Jian Cao, Guandong Xu, and Yudong Tan. 2021. TFROM: A Two-sided Fairness-Aware Recommendation Model for Both Customers and Providers. arXiv preprint arXiv:2104.09024 (2021).

[91] imperva. Anontmization. Retrieved February 6, 2023. from https://www.imperva.com/learn/data-security/anonymization/

[92] Yamazoe, T., Omote, K. (2020). Safety Analysis of High-Dimensional Anonymized Data from Multiple Perspectives. In: Kutyłowski, M., Zhang, J., Chen, C. (eds) Network and System Security. NSS 2020. Lecture Notes in Computer Science(), vol 12570. Springer, Cham. https://doi.org/10.1007/978-3-030-65745-1_6

[93] Shakil et al. (2022). Towards a Two-Tier Architecture for Privacy-Enabled Recommender Systems (PeRS). In: Wang, G., Choo, KK.R., Ko, R.K.L., Xu, Y., Crispo, B. (eds) Ubiquitous Security. UbiSec 2021. Communications in Computer and Information Science, vol 1557. Springer, Singapore. https://doi.org/10.1007/978-981-19-0468-4_20

[94] Jing Qin. 2021. A Survey of Long-Tail Item Recommendation Methods. Wireless Communications and Mobile Computing. vol. 2021. Article ID 7536316, 14 pages. https://doi.org/10.1155/2021/7536316

[95] Moez Ali, Medium. Top Explainable AI (XAI) Python Frameworks. 2022. Retrieved February 6, 2023. from https://moez-62905.medium.com/top-explainable-ai-xai-python-frameworks-in-2022-94ff4610b0f5

[96] Yassine Himeur, Aya Sayed, Abdullah Alsalemi, Faycal Bensaali, Abbes Amira, Iraklis Varlamis, Magdalini Eirinaki, Christos Sardianos, George Dimitrakopoulos, Blockchain-based recommender systems: Applications, challenges and future opportunities, Computer Science Review, Volume 43, 2022, 100439, ISSN 1574-0137, https://doi.org/10.1016/j.cosrev.2021.100439.

[97] Kelvin, Towards Data Science. Introduction to Federated Learning and Challenges. 2020 Retrieved February 6, 2023. from https://towardsdatascience.com/introduction-to-federated-learning-and-challenges-ea7e02f260ca

[98] Silvia Chiappa, William Isaac, DeepMind. Causal Bayesian Networks: A flexible tool to enable fairer machine learning. 2019. Retrieved February 6, 2023. from https://www.deepmind.com/blog/causal-bayesian-

networks-a-flexible-tool-to-enable-fairer-machine-learning

[99] Hayes, C.F., Rădulescu, R., Bargiacchi, E. et al. A practical guide to multi-objective reinforcement learning and planning. Auton Agent Multi-Agent Syst 36, 26 (2022). https://doi.org/10.1007/s10458-022-09552-y

[100] Roy, D., Dutta, M. A systematic review and research perspective on recommender systems. J Big Data 9, 59 (2022). https://doi.org/10.1186/s40537-022-00592-5

[101] *Qian Li, Xiangmeng Wang, Zhichao Wang, and Guandong Xu. 2023. Be Causal: De-Biasing Social Network Confounding in Recommendation. ACM Trans. Knowl. Discov. Data 17, 1, Article 14 (January 2023), 23 pages. https://doi.org/10.1145/3533725*

[102] Chang W, Sun D, Du Q. Intelligent Sensors for POI Recommendation Model Using Deep Learning in Location-Based Social Network Big Data. *Sensors*. 2023; 23(2):850. https://doi.org/10.3390/s23020850

[103] F. Ji, J. Wu, F. Chiclana, S. Wang, H. Fujita and E. Herrera-Viedma. (2023) The Overlapping Community-Driven Feedback Mechanism to Support Consensus in Social Network Group Decision Making. in *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 9, pp. 3025-3039, Sept. 2023, doi: 10.1109/TFUZZ.2023.3241062.

[104] Wang, Y., Dai, Z., Cao, J. *et al.* Intra- and inter-association attention network-enhanced policy learning for social group recommendation. *World Wide Web* **26**, 71–94 (2023). https://doi.org/10.1007/s11280-022-01035-0

[105] Perifanis, V., Drosatos, G., Stamatelatos, G., & Efraimidis, P. S. (2023). FedPOIRec: Privacy-preserving federated poi recommendation with social influence. *Information Sciences*, *623*, 767-790.

[106] Dharmendra Kumar Singh Singh, Nithya N., Rahunathan L., Preyal Sanghavi, Ravirajsinh Sajubha Vaghela, Poongodi Manoharan, Mounir Hamdi, Godwin Brown Tunze. (2022).Social Network Analysis for Precise Friend Suggestion for Twitter by Associating Multiple Networks Using ML. International Journal of Information Technology and Web Engineering (IJITWE) 17(1). DOI: 10.4018/IJITWE.304050

[107] Liu, H., Zheng, C., Li, D., Zhang, Z., Lin, K., Shen, X., Xiong, N.N. and Wang, J., 2022. Multi-perspective social recommendation method with graph representation learning. *Neurocomputing*, *468*, pp.469-481.

[108] Liao, J., Zhou, W., Luo, F., Wen, J., Gao, M., Li, X., & Zeng, J. (2022). SocialLGN: Light graph convolution network for social recommendation. *Information Sciences*, *589*, 595-607.

[109] Hui, B., Zhang, L., Zhou, X. *et al.* Personalized recommendation system based on knowledge embedding and historical behavior. *Appl Intell* **52**, 954–966 (2022). https://doi.org/10.1007/s10489-021-02363-w

[110] Kitsios, F., Mitsopoulou, E., Moustaka, E., & Kamariotou, M. (2022). User-Generated Content behavior and digital tourism services: A SEM-neural network model for information trust in social networking sites. *International Journal of Information Management Data Insights*, *2*(1), 100056.

[111] *Zhiwei Liu, Liangwei Yang, Ziwei Fan, Hao Peng, and Philip S. Yu. 2022. Federated Social Recommendation with Graph Neural Network. ACM Trans. Intell. Syst. Technol. 13, 4, Article 55 (August 2022), 24 pages. https://doi.org/10.1145/3501815*

[112] Park, D., Kim, S. H., & Elmqvist, N. (2023). Gatherplot: A Non-Overlapping Scatterplot. *arXiv preprint arXiv:2301.10843*.

[113] Zhang, L., Zhou, X., & Shen, Z. (2023). Multimodal Pre-training Framework for Sequential Recommendation via Contrastive Learning. *arXiv preprint arXiv:2303.11879*.

[114] Chatti, M. A., Guesmi, M., & Muslim, A. (2023). Visualization for Recommendation Explainability: A Survey and New Perspectives. *arXiv preprint arXiv:2305.11755*.

[115] Raval, A. S., & Borisagar, K. R. Document Recommendation: An anatomization of neoteric advancements, problems and potential scope.

Author Biography:

Ms. Aanal Raval is a full-time Ph.D. scholar at Gujarat Technological University – Graduate School of Engineering and Technology. She has done her B.E (in 2018) and M.E. (in 2020) from

Silver Oak College of Engineering and Technology (GTU affiliated till 2020). She has done her dissertation work in Machine Learning. She has published more than 8 journal papers. Her area of interest is Artificial Intelligence, Machine learning, Natural Language Processing, Deep learning, Reinforcement Learning and Computer Vision.

Dr. Komal Borisagar- She completed Under-Graduation in Electronics & Communication branch from C.U.Shah College of Engineering and Technology - Wadhwan under Saurastra University.in 2002, Post-Graduation in in Electronics and Communication (Communication System Engineering) from the Charoter Institute of Technology, Changa under Gujarat University in 2008, and Doctor of Philosophy in Electronics and Communication Engineering from Shri Jagdish Prasad JhabarmalTibrewala University, Vidyanagari, Jhunjhunu. She has published more than 66 research papers in various scopus, IEEE indexed international conferences and journals in the area of image processing and communication engineering.