Document Recommendation: An anatomization of neoteric advancements, problems and potential scope

Aanal S Raval*
Research Scholar, Gujarat Technological University, Ahmedabad, Gujarat
email id: aanalraval@gtu.edu.in
ORCID: 0000-0003-3848-0882

Dr.Komal R. Borisagar
Associate Professor, Gujarat Technological University, Ahmedabad, Gujarat
email id: asso_komal_borisagar@gtu.edu.in
ORCID: 0000-0002-1047-6091

Abstract
Accompanying the increased use and storage of digital documentation, like either a repository of research articles or legal and medical documents, finding the proper article is a dullsville task. With the increased number of articles, searching for pertinent documentation from constantly growing repositories becomes a sluggish and incommodious chore. In recent years, various evaluation datasets, solutions and methodologies have been proposed claiming considerable progress with state-of-art in area of recommender systems. In this study, we present the scope of document recommendation whether in academic paper or patent citation, assisting decisions in legal/ medical decisions, news or relevant content finding for digital services. Then we depict the present methodologies and approaches along with datasets. Analysing various aspects, we show the differences. Furthermore, we unveil the challenges and directions to work on them. This survey can help enthusiasts and researchers of this field in comprehending recent trends, by providing with cutting edge knowledge.
Keywords: Recommender systems, Text recommendations, Document recommendation aspects and goals, ML/DL/ RL methods for document recommendation

1. Introduction:

The evolution of the information era, sets in motion an opportunity for various online publications such as research articles, legal documents including precedents, codes, constitutions, rules, regulations, acts and laws, medical documents including various reports and drug information sources. With rapidly growing works of literature and articles, it is an eminently tough task to find a suitable and propitious article. This work is crucial not only for researchers, biomedical data curators, and legal document analyzers like judgments and related legal issues but also for matching an individual’s profile with job descriptions appropriately in job searching. Although some academic search engines and various digital repositories can manage and search pertinent papers according to researchers, but in turn to do that, individuals must properly express the keywords and smartly filter the relevant ones from the list. But this thing is quite difficult for beginners. So subjugating above problems, document recommendation helps in finding relevant articles from a massive repository.

Many existing surveys emphasis on recommendation in either article citation or finding relevant articles or news stories. Some works [56] concern on paper recommendations considering characteristics like context of content, some analyse recognition of context of citations, classifying them based on popularity, pinpointing citation behaviour and examining task specific dimensions to compare the approaches to citation estimation [57]. In addition, works focusing on news recommendation [29], has shown challenge of quality content and effects of content on users, which in general is applicable to readers of every kind of content or articles. Analysis of related work and methodologies shoe that they don’t focus on effects of image aspects. We analyse the point of considering image as feature along with attention worthy points like privacy and decentralized computations. Apart from these, considering explainability, fairness, multimodal and long tail in recommendations, we provide insights to possible solutions to all these. This work reviews different document recommendation systems and analyses them illuminating unrealized and dormant new directions. Examining state of art techniques and results, this survey can assist in the research and evolution of futuristic and trailblazing developments.

The organization of this paper is as follows: Section 2 shows the background, general characteristics and motivation behind document recommender systems. Section 3 conveys the detailed categorization along with comparison of research in document recommendation task. Section 4 and 5 analysis datasets and use of evaluations metric according to task specifications. Section 6 is analysis part from the research exploration conducted. Section 7 includes detailed challenges currently faced or where the scope of work exists. Section 8 explains the insights to techniques that can be employed for new future researches.

2. Current scenario of Document Recommendation:

A large number of readings materials including technical, non-technical, legal, medical documents, news articles and research papers are published every day. Consideration of such voluminous data, its continuous spread and unstructured format, are completely dependent on factors like its recency and popularity. For reading purpose, signing in or creating profiles is generally not required, which means users rarely give any feedback, So in this case the consumption behaviour can be totally analysed by click logs, tracking a session or through browser cookies. This motivated the need of creation of user profiles. The foremost need of user profile to recommend is to have some historical data to start with. But if a new user or a new item enters, either a small data or no data will be available. So along with user profile creation, need to tackle the problems of cold start and data sparsity too arose.

Apart from these factors, user preferences change from time to time revealing some to be short term as well as long term preferences. In addition item churn too plays an important role here. Document or reading recommendation are prone to over specialization and filter bubbles where only certain entities are recommended instead of diverse and up-to-date content suggestions. Thus narrowing users' prospect and choice by depriving their exposure to new content, which can actually be in demand, despite of its low popularity can arise need to handle long tails.

3. Research in Document Recommendation:

In this section, we present the existing research with its analysis according to various dimensions classifying the task of document recommendation to different sub categories. Based on type of textual bulk literature has been classified into seven categories. The given figure-1 describes the categories of application.

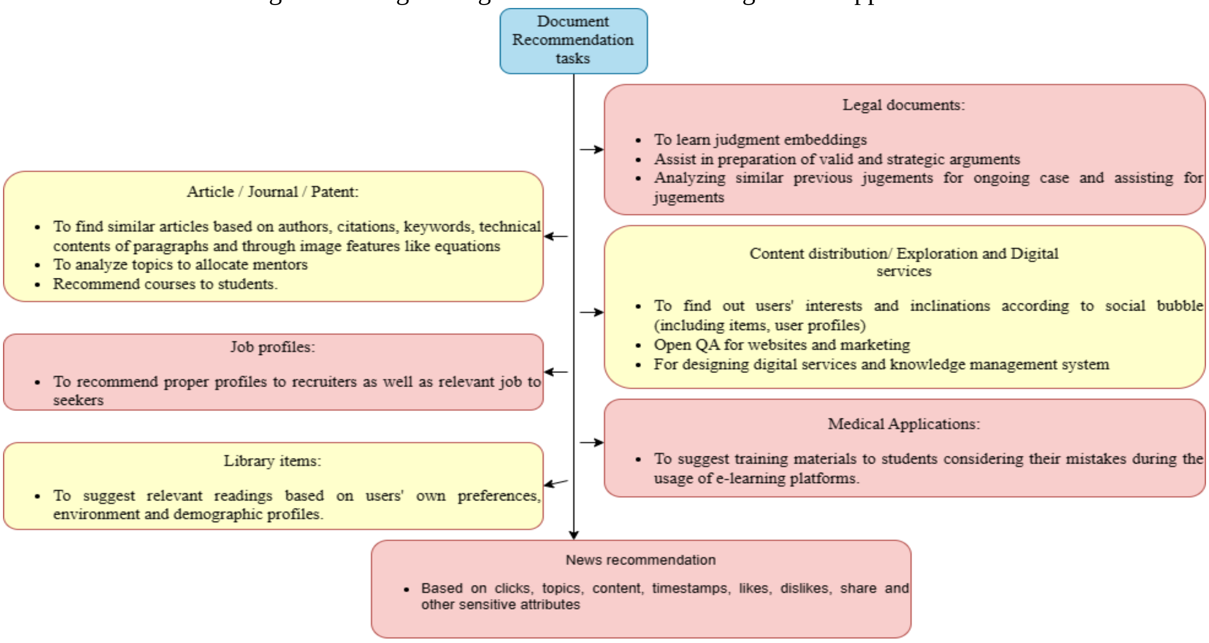

Figure-1 Categories of Document Recommendation applications

3.1 Recommendations for Legal documents:

Recommendations in legal domain include purpose of analysing previous judgements assisting current one, to craft and define set of rules, updating or deleting rules required as per analysed scenario and consider relevant judgements for the ongoing one.

Making use of pre-learned word embedding on Legal domain-specific that possessed legal semantic knowledge, [1] proposed a pre-learned embedding framework P-LDRS that learned Doc2vec embedding. Judgement embeddings can also be learnt through this framework which is distributed MapReduce and Spark like computing node clusters to address the issue of scalability. The proposed framework outperforms traditional doc2vec with accuracy of 88%. [2], in order to remove the noisy data and reduce the size of corpus, proposed a framework to keep relevant dictionary words in corpus discarding the noisy elements, which in turn can improve efficiency of Doc2vec. As with time, the number of judgements increases quadratically [73] Considered relevant judgements from every clusters using pairwise similarity scores. Given table 1 shows summary of Recommendations for Legal documents

Table-1 Features used and goals of Recommendations for Legal documents:

Paper	Model	Features used	Goal to address
[1]	Law2Vec, LeGlove, Legal Word2Vec	arguments, facts, issues, verdicts of the court, judge name, time-based information,	To craft context rules by training legal documents on Skipgram, CBOW and Glove
[2]	Doc2vec	Judgement embeddings	To deal with noisy corpus for time and space efficiency
[73]	Doc2vec, Clustering, pairwise similarity scores	Relevant Judgement embeddings	Scalability issue

3.2 Recommendations for Journal / Technical or Non-technical Article/ Patent / citation:

This section explores the recommendations in technical non-technical articles or research papers considering the features like summary of content, citations, title, abstract, user profiles and connections, feedback and other semantics.

Based on ThaiJO (TCI Thai Journals Online Database), [3] Developed a recommender system by tf-idf and cosine similarity scores to recommend most suitable article. [4] used tf-idf along with word2vec to weight its vectors. [6] Proposed a two-stage deep learning model to select the textual information, cooperative patent classification embedding and re-ranking of candidates using citation information. Cooperative patent classification embedding was generated using similarity in patent text and citation. For this task, they synthesized a dataset called PatentNet summarizing citation information of examiner, citation dependent on textual data and metadata of about considering 110,000 number of patents. This model outperformed existing ones with mean reciprocal rank of 0.2056. To recommend articles considering contextual and semantic features, [7] used doc2vec to generate embeddings of semantics and features, and XGBoost to show the impact of different attributes like keywords, abstract and title on publication with 84.24% accuracy. [8] Proposed transfer learning-based model by fine-tuning pre-trained contrastive learning framework- distilled version of RoBERTa to learn document representations. And then these features were trained for downstream task of classification using paper attributes like title, keywords, aim, abstract and scope. Results with data consisting of 414512 papers, 331464 for training and 83048 for testing showed accuracy of 94.96% for top 10 recommendations. [9] Proposed a model with three stages: In first, a GRU for word embeddings with contexts, Secondly, high way networks to combine different features like word with document level resulting in knowledge of linguistic structure levels and thirdly two layers of CNN to detect the patterns of semantics. Through semantics representations and metadata about users' profile, recommendations were made. Metadata along with user-item rate was predicted with help of multitask learning. [10] Developed a system to recommend reference that is missing from local and global context. For this task, they first created an embedding by hierarchical attention network, filtered out top k predictions using nearest neighbour and finally re-ranked it using SciBERT. [11] Used hierarchical agglomerative text clustering and Principal Components Analysis for grouping the similarities in documents. [12] Used the idea of representation creation with the help of mappings and linear concept approximation for alignment of monolingual representation for recommending multilingual documents. [16] In order to unveil semantic structure from documents used Latent Dirichlet Allocation. [19] Used search intent, users access log, tf-idf and text mining techniques to determine reason due to why user reads documents instead of concerning who reads. [20] Utilized tf-idf with K-means, cosine similarity and agglomerative clustering to extract relevant articles based on individuals' reading. [22] Proposed LDA based method to recommend articles using Jensen-Shannon distance and Kullback-Leibler divergence. [23] too used LDA along with bigram model and Lemmatization. [27] proposed some indirect features and their extraction methods for given research article like Keyword Diversity Measurement (KDM), Sentence Complexity Analysis (SCA), Citation Analysis Over Time (CAOT), Scientific Quality Measurement (SQM). [28] Used Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA) for document recommendation. [29] Constructed a heterogeneous relation graph to extract multiple objects and their respective relations through TF-IDF and adjacency matrix. Then relations between users and papers were set with personalized learned weights. Then based on recommendation scores, these features, and historical preferences, top N recommendation list was generated. [62] Proposed the solution to diverse personalized learning with heterogeneous network and personalized weight learning. [64] Proposed one shot learning and attention based CNN and Bayesian learning approach for both personalized and non personalized setting. Here results lowered on personalized setting which concludes its inappropriateness for online environment. [65] Foremost constructed a citation network representing papers and authors as vertex, then fragments are extracted for each citation link explaining about its information. And finally recommend papers through vector similarity, obtained through network similarity learning. [70] Argued consideration of author, co-author relations with title. They developed a reinforcement based network embedding to enhance embeddings of relation between co authorship and textual information, to address cold start problem. Given table 2 shows summary of Recommendations for Journal / Technical-Non-technical Article/ Patent / citation.

Table-2 Features and goals of Recommendations for Journal / Technical-Non-technical Article/ Patent / citation:

Paper	Model	Features used	Goal to address
[3]	Tf-idf, cosine similarity with ranking	Abstract summary, Thai with English keywords	Journal recommendation
[4]	weighted Word2vec with TFIDF	Non semantic features of text	Classification of article text
[9]	GRU with CNN	User profile and various document features	Recommend paper based on semantic features
[8]	Transfer learning with language model (LM)	title, keywords, aim, abstract and scope	Top 10 recommendations
[10]	Hierarchical network, KNN, LM	Document embeddings	recommend reference from local and global context
[11], [20], [21]	Cosine similarity, k-means, agglomerative clustering	Content semantics	Grouping of similar documents
[12]	Mappings and linear concept approximation	Document embeddings	Multi lingual document recommendation
[19]	Text mining techniques	Search intent, user access log	To determine user interest
[22], [23], [27], [28]	Latent Dirichlet Allocation with similarity distance, LSA and NLP	Keyword Diversity Measurement, Sentence Complexity Analysis, Citation Analysis Over Time, Scientific Quality Measurement	Recommend articles
[29]	Heterogenous relation graph with tf-idf and adjacency matrix	recommendation scores, weights of user-paper, historical preferences	Top-N recommendations
[62]	heterogeneous network with personalized weight learning of users.	Users' interconnectedness, users' personalized weights, sum of products of the recommendation score of the candidate paper with the target user's personalized weight	recommend diverse personalized papers
[64]	One shot learning and attention based CNN	Uncertainties in user preference, feedback	personalized and non personalized setting
[65]	Citation information network	Citation, its explanation	Citation analysis
[70]	Reinforcement based network embeddings	Relations between Co-authors and text	Citation prediction, cold start issue

3.3 Recommendations for Text bulk in Content distribution/ Exploration and Digital services:

Recommendations in this section considers the online available textual bulk user history and profile details, proposals, feedbacks in form of QA, quotations and data from e-commerce sites as document (textual bulk). Analysis of these data can reveal appropriate information from chats and question-answers, help e-commerce sites for analysis and appropriate distribution of content after knowing its semantics, explore profiles and analyse them to help for better personalized recommendations, proper investigation and tackling of proposals or quotations of business platforms as well as proper communication and process flow among suppliers, intermediaries and consumers on platforms like autoDX.

[15] Proposed attention based mechanism called SEAN (social explorative attention network) to encode an incoming document to user dependent contextualized vectors for related words and sentences in new document. To extend the scope of users or to tackle with new users or users with less reading history, variants of given mechanism, SEAN-END2END- to combine any users' vectors with their friends' context vectors and SEAN-KEYWORD for users' interesting words from document and history. For contextual vectors, friends selection is done through experimenting Monte Carlo Tree Search and ϵ -Greedy. For some searching criteria for proposal creation, in knowledge management system in sales, [21] developed similarity based recommendation for proposals. For this task, the features were extracted for model based on industrial context, then word embedding models were applied, results of both were considered, and finally according to similarity score and date, they were ranked to recommend. [25] Deployed LSTM for preprocessing and corpus learning, following LSTM with graph embeddings for topic recommendation based on interest and content recommendation based on semantic annotation. [26] Modeled QA documents for websites using content based recommendation for CB score, Collaborative filtering recommendations for finding the user similarities, Complementarity based recommendation for classification of complementarity relations. These outputs were combined for lists and rankings using hybrid recommendations. [14] Introduced content based recommendation for the quotation task by ranking the paragraph contents provided by authors and then scoring the span of the given document according to predicted quotations and relevance. For both the tasks, SoTA BERT-based models were used and were combined with open-question answering to generate recommendations. [67] Proposed deep learning based approach with two levels, first with bidirectional encoder to represent different textual features and second sequential recommendation model based on attention to derive item embeddings for e commerce platform analysis. [68] Performed document ranking based on two kinds of features first considering query and signals like click-through rate and sessions, second based on users' interest, click history and preferences. In order to solve the appropriateness of comments or shared knowledge on stack overflow, [71] proposed some features (degree of relatedness among text/ comment, Author-based, Temporal, Text similarity and semantic features) to evaluate pairwise connectedness of comments text through random forest (balancing with SMOTE) and then creating its clusters of text for same comment along with its relatedness with other clusters. Features were created using NLP techniques such as cosine similarity, jaccard similarity glove_cosine and weighted_glove_cosine. Given table 3 shows summary of Recommendations for Content distribution/ Exploration and Digital services.

Table-3 Features used and goals of Recommendations for Content distribution/ Exploration and Digital services:

Paper	Model	Features used	Goal to address
[14]	Content based recommendation with BERT models	Paragraph vectore, document span, relevance, open QA	BQuotations
[15]	social explorative attention network	users' vectors, their friends' context vectors	Creation of user dependent contextualized vectors
[21]	Similarity distance and embeddings	Embedding vector features, similarity score, date	Proposal creation
[25]	LSTM with graph embeddings	User interest, content semantics	Content recommendation with semantic annotation
[26]	Content based and Collaborative filtering with hybrid recommendations	QA, CB scores	Complementarity relations among users for content distribution
[71]	Random Forest with clustering	Relationships among authors and semantics of text, time elapsed, similarity in comments, degree of relatedness (Unrelated, Slightly Related and Strongly Related)	To provide appropriate knowledge from stack overflow's comments and text

3.4 Recommendation for Job profiles:

Concerning job profile, analysis of text can help both job seekers and recruiters to find appropriate job according to preferences and to recruit proper candidates according to their background, knowledge analysing other factors respectively.

For job recommendations, [17] applied an approach with implicit skills extracted through NLP techniques, after a document similarity match and projecting them to respective semantic space after applying Doc2vec. They define implicit skill as those which are not mentioned in resume but can be present in context of some role, industry or geography. Thereby, several skills were considered to match candidates' CVs to job descriptions. [63] Employing various NLP techniques such as NLP-token count matrix, percentage by cosine similarity, phrase matching, spacy for features creation and matching, calculated job satisfaction and retention parameters, used linear and ridge regression. Parameters were calculated experimenting various techniqes like Naïve Bayes, Support Vector Classifier, K nearest neighbour, decision tree, random forest, and Multilayer perceptron. Given table 4 shows summary of Recommendations for Job Profiles.

Table-4 Features used and goals of Recommendation for Job profiles:

Paper	Model	Features used	Goal to address
[17]	Doc2vec with the semantics of resume	Explicit skills mentioned, implicit skills : type of role, industry or geography	To match the candidate's CV for specific job
[63]	Various traditional ML techniques	Job satisfaction, retention	To recommend job to a candidate

3.5 Recommendations for Library items and Academics:

In this category, considering user factors like demography, history, environment of user, semantics of book or course and academic history of user, book, materials and online courses can be suggested.

Using an Artificial Intelligence Recommendation, [5] proposed a Knowledge-based Environment architecture for Personalized Learning (KEPLAIR) to find relevant reading recommendations for library users. It consists of three main modules, first, a Learning Manager to supervise interactions with humans, recommend the learnt objects, in building learners as well as in checking their performance. Second, the social manager records the log of inert user communications and their individual behaviour. And a harvesting manager, to store continuously updating metadata of learning objects. Its knowledge base focuses on four kinds of information, interest of users' goals, demographic profile of users, learnt objects from personal preferences, and environment of the learner. Their experiments show that when it contained prior information about users' interests, it was able to predict greater than 2/3 of recommendations. [13] For the task of document collection, applied graph-based ranking on content-based recommender systems generating index from the abstract. [18] Developed a methodology to compare the topic proposals in order to allocate a mentor. For this purpose, they used Rapid Automatic Keyword Extraction to produce keywords and documents were matched using Leveshtein distance similarity. Many other works [59], [60] observed students' performance for examinations and their advisory system to choose courses based on their current performance on various subjects. [72] Figured out the problem of discussion forums and chats in Massive Open Online Courses (MOOCs), where sometimes answering to each and every person becomes much more inconvenient for the expert. For this purpose, they propose a system to analyze students' chats to recommend with proper learning materials, considering its sentiment, intent, level of concepts cleared or confusion in it. Given table 5 shows summary of Recommendations for Library items and Academics.

Table-5 Features used and goals of Recommendations for Library items and Academics:

Paper	Model	Features used	Goal to address
[5]	KEPLAIR	users' goals, demographic profile of users, learnt objects from personal preferences, and environment of the learner	Library recommendations
[59], [60]	Euclidian distance, KNN, decision making	Subjects	Students' behavior prediction, student support advisory
[13]	Content based system with graph based ranking	abstract	Recommend thesis from e library, to reduce

			search space of document similarities
[18]	Leveshtein distance	Keywords	To allocate mentor based on topic
[72]	Rule based recommendations	Details of topic for specific course, sentiment, intent, level of concepts cleared or confusion from chat texts	To recommend students with suitable learning materials.

3.6 Recommendations for documents in Medical field:

Analysing medical QA documents based on chats, doctor's and patient's background etc., assistance could be provided for diagnosis with a common set of questions to be asked or common set of symptoms to be recognized. Analysing these data, some general tasks could be automated like preventive measures, drugs and relevant needed resources can be recommended.

In order to suggest subject matter and some relevant stuff to students as a part of training while using e- learning platforms i.e., an interaction with virtual patient [24] experimented with Learning to Rank approaches like Lambda Rank, MART, ListNet, RankNet, LambdaMART, RankBoost, Random Forest (RF) and Coordinate Ascent. Their results show RF outperformed all with Normalized discounted cumulative gain (NDCG) of 0.89, when no new documents were present. [58] Employed fuzzy inference rule and decision tree rule based system for recommendation of medicines and preventive measures for diabetes. Recommendations made include preventive measures or care to lead a natural life, exercises, nutrition and other required medications. [62] Experimented various ensemble learning methods as well as traditional techniques like Multinomial Naïve Bayes, logistic regression, and stochastic gradient descent along with embeddings and clustering methods like SMOTE and t-SNE. Clustering methods were applied for balancing datasets. The task addressed here is to recommend drugs based on their reviews. [69] Developed a system to assist Online healthcare communities with recommendations of needed steps to be taken based on patients' need, contents of QA and Doctor' s background information Initially creates an embeddings of information about patient's need, doctor's background and respective utterances. Then they applied attention based mechanism to capture the interaction among those to get the relevance score. And in last stage, they deployed Gated Recurrent Unit (GRU) with single NN to match generated scores and generate a ranking list based on it. Given table 6 shows summary of Recommendations for Medical field.

Table-6 Features used and goals of Recommendations for Library items and Academics

Paper	Model	Features used	Goal to address
[24]	Learning to rank approaches	Interactions between patients and students while diagnosing virtually	To address with medical needs of patients that are unintentionally missed out
[58]	Decision tree and fuzzy rule based	Risk or stage of diabetes, attributes of PIMA dataset	To recommend medicines and preventive measures
[62]	Various ensemble learning and traditional ML approaches	Reviews/ Feedback	To recommend drugs
[69]	Attention based mechanism, GRU	QA of multiple rounds	Help Online healthcare communities with recommendations to tackle with resources like time and work pressure

3.7 Recommendations for News:

Analysis of text becomes important in news recommendations, for user modelling and user engagement prediction based on semantics of news articles, references made, history, trendiness of topic, clicks title, time spent on link and with other user sensitive attributes like gender, age, profession, demography etc.

[29] Explored research in news recommender including traditional approaches like collaborative filtering (CF), content-based filtering (CBF) and hybrid, factorization models like matrix factorization (MF), Non-negative matrix factorization (NMF), Tensor factorization, Probabilistic matrix factorization (PMF), Bayesian personalized ranking (BPR) and Generalized linear modeling (GLM) along with neural extensions of these techniques. In addition, this work summarized some libraries and platforms like Apache Mahout (ML library in java containing CF algorithms for real world datasets), Idomaar (framework enabling evaluation in recommender systems with flexibility of programming languages), StreamingRec (includes various news recommendation techniques), CLEF NEWSREEL and Open Recommendation Platform (ORP) CLEF NEWSREEL platform (using which new recommendations can be developed with plista dataset) and Hugging Face libraries. Furthermore, they discussed the DL based models for recommendation, 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). [47] surveyed methods like Entity centric (latent representation of user and news profile with vector space model), Path based methods (showed user-item, item-item similarity from knowledge base) and neural-based methods (creating latent representation with NN and combining their encoded information (embeddings) by knowledge graphs). [30] Experimented context based features such as title, popularity, freshness and click on various existing models. [44] used hierarchical attention for users' subtopic interest, after which they combined interests of subtopics to level of topic for user interest. [45] In order to model users, used attention network which is candidate aware for the task of selecting clicked news considering its relatedness from candidate one. Following table summarizes new recommendation tasks with features based on applied models. [48] Deployed hyper graph with affinity matrix, ranking and query vector considering timestamps, weights of stakeholders and coverage weights of authors. To avoid Over-specialization and under-representation of popular or unpopular authors, weights were measured by the popularity of authors, a graph was constructed depicting scaling of weights as per user. Then these scores were retrieved, normalized and product of coverage weights of authors with relevance scores was taken to evaluate weight of query vector. Diversity of topics and keywords were taken care of through relevance scores based on topics which aren't recommended yet. Given table 7 shows summary of Recommendations for News.

Table-7 Features used and goals of Recommendations for News:

Paper	Model	Features used	Goal to address
[31]	Attention with Transformers (NM/ UM)	Click, non-click, quick close, share, finish, dislike	User engagement prediction
[32]	Attention with Transformers	sensitive attributes (gender, age, profession)	Role of sensitive attributes for recommendation
[33]	Graphical embeddings (GNN) with attention (NM/ UM)	Click	Representation of user based as entity
[34]	hard similarity with 3 D CNN	Click, Title	User modeling
[35]	Attention with type of event classification (UM), LSTM with attention and node2vec (NM)	Title, User – news graph, click	Representation of user history for recommendation
[36]	Combination of the transformer with Pretrained Language Model (UM), Attention (NM)	Click, title	Text understanding for recommendation
[37]	Encoder (user modelling) and CNN with trandformer, co-attention, graph co-attention (news modeling)	Title, entity, click	to learn candidate aware user interest representation and vice versa

[38]	Transformer	Click, title	To study task of sequential recommendation for news
[39]	LSTM with attention and co-attention (NM), GCN with attention (UM)	Click, title, summary abstract	To learn variety in interests and behaviour of users
[40]	LSTM, attention, deep walk	Click, non-click, follow, like, share, comment, heterogeneous graph embedding, user embeddings, keywords, label/ tag, category, coritivity score	To analyse variety in user behaviour for higher order representation of features to tackle predict behavious of new users.
[41]	Memory-reasoning network	Click	User modeling
[42]	Gated Recurrent Network with ID embedding	Click, customer id	User modeling
[43]	Content-popularity joint attention (UM), self attention, co-attention, gating, (NM) Bayesian Personalized Ranking (BPR) (train clicks and non clicks)	Click through rate (CTR), title, recency, entity,	To predict recency popularity of news (user interest modeling)
[44]	Transformer with attention (NM), hierarchical attention (UM)	Click, Title, entity	To model relations between entities in texts and context from titles
[45], [46]	ViLBERT	Click, Image, title	To capture multimodal features
[48]	Hypergraph	Diverse scores of topic, timestamps, stakeholder weights, relevance scores of authors for each user	To solve popularity bias and introduce fairness
[66]	NN with KG and RL	Latent representation of article (AG), interaction among Anchor graph, reasoning	Diverse and fair news recommendation

4. Datasets:

As long as the entity to be pondered in either case of recommendations studied is bulky text, textual document data are here analysed. Here we discuss all kinds of datasets that are either publicly available or synthesized or crawled.

FullTextPeerRead is a context aware dataset containing cited references and metadata of paper. Identical to FullTextPeerRead, ANN dataset too has attributes like id’s of target and source and extra texts of citations. ACL-200 has papers with contexts. RefSeer dataset has been released under Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. This consists of 3 tables: paper (with id and respective cluster), citations (citation id, cluster and paper id), and citation index (citation id and context) [76]. The arXiv dataset provides open access to scholarly articles around 1.7 million from various disciplines of computer engineering as well as from multiple disciplines like statistics, maths, economics, quantitative biology and electrical engineering. This dataset provides relevant features like title, categories, authors, abstracts, full text extending scope to various tasks like paper recommendation, trend analysis, co-citation networks, interfaces of semantic search, category prediction and knowledge graph construction [77]. HepTh is a graph network dataset for showing the relations among authors and respective citations.

Aminer is a non-commercial dataset showing relation between researchers and respective statistical correlations. Aside from indexing 130,000,000 researchers conjointly else exceeding 265 million publications, this dataset is used to analyze social networks assisting in recognition of topic modelling, academic performance, to find expert, reviewing recommendations, course search, geographic search and association search [79]. DBLP dataset is a citation network dataset from DBLP, ACM and Microsoft Academic Graph, containing more than 632,752 citations and 629,814 papers, where each paper has its own title, abstract, authors, venue and year. Apart from topic modelling, clustering of network with co-information, uniqueness here is that it has analysis of influence of each other in citation network and can find papers that are most influencing [80]. RARD (Related-Article Recommendation Dataset) is a dataset from Mr. DLib and digital library Sowiport. This includes approaches to recommendation, types of features and soyrce od its extraction, time logs and click logs [82]. As compared to RARD, its successor second version has 187% more features, 64% extra recommendations, 50% excess clicks and 140% additional documents.

Drug review dataset has reviews on particular drug along with customer's side rating. JRC-Acquis is a multilingual corpus from European Union for legal domain [78]. Similar skills and Skill2vec contains a large number of more than 1,400,000 job descriptions, where skills are extracted and a list is created of those with similar context (includes skills for specific job description) [81]. Wikipedia Graph and Related Entity Recommendation Dataset contains normalized detailed Wikipedia hyperlink entity graph considering many languages each with different features. It contains query entities with their respective relevant entities which can be further used as ground truth as well as evaluation purpose. Training of entity embeddings has been done through lg2vec. Yahoo Search Query Log To Entities is benchmark to link the entities. Yahoo Answers Novelty Based Answer Ranking: This is a medical dataset where relevant target question context and answers are annotated to mapping of textual assertions.

Roularta dataset is collected from a Belgian multimedia group consisting of weekly/ monthly news magazines on business trends, sports, women's interests and lifestyle, local, medical and other professions. This consists of summary of content, title and information from other articles. It has IPTC (International Press Telecommunications Council) tags for bartering of news among various agencies. For user interactions, the popularity of authors have skewed distributions. The task of generation of embeddings and feature extraction of text has been accomplished through CNN. This dataset can reveal information about short head and long tail authors. Plista dataset is comprised of news data collected from 13 portals, including statistical analysis of CTR, publishers' create and updates, and impression showing bias in popularity and sparsity, along with sum, mean and standard deviation of mentioned features. [50] Yahoo Webscope is a reference library of datasets for scientific experiments and academic use on free basis. Its benchmark datasets include Front Page Today Module User Click Log Dataset, Front Page Today Module User Click Log Dataset and Yahoo News Video dataset. The Yahoo News Annotated Comments Corpus dataset is dataset of comments and threads in response to news articles. Yahoo News Ranked Multi-label Corpus provides actual text from Newsroom, Yahoo news, Tumblr and many other for multilabel tagging and ranking enabling researchers to use their own features. Yahoo! News Sessions Content provides user session wise information like clicks, number of news articles, tokens, publication details, timestamps and entities (such as organization, location, person). One of pros of this dataset is that it contains anonymised user and news data so that identity isn't revealed.

Adressa, For user interactions, the popularity of authors have skewed distributions. Considering 11207 articles and 561733 users, this dataset reveals news connection with anonymized users. This dataset contains information like subscribers-nonsubscribers, session identification, click counts, reading times, title, author, category, keywords, entities and body. [49]. Hacker news is a dataset of stories and comments from Hacker news containing story id, author, time at which it was written and its popularity as features. This MIT Licensed dataset enables users to share news, IT projects demonstration, posting jobs asking questions and commenting on stories. BuzzFeed news provides dataset for news stories, entertainment along with fake news, patterns in social media and news.

MIcrosoft news dataset (MIND), In area of news recommendation, this large-scale dataset is a benchmark. It contains about 160k news articles, and more than 15 million impression logs with 1 million users. History of click behaviours, click events and non-clicked events are present for each log. For every article features considered are title, body, abstract, category and entities. Concerning user privacy, behaviour logs of users are anonymized. Fake news provides two datasets, first FakeNewsAMT collected by combination of crowdsourced and manual annotations, and second Celebrity collected through web scrapping. FakeNewsAMT includes news about education, technology, sports, business, entertainment and politics. Celebrity dataset contains stories on socialites, politicians, singers and actors. So for celebrity dataset, legitimate new were extracted from fashion-style and entertainment section, while fake new category included online gossip stories.

Apart from these, datasets like Global Database of Events, Language and Tone (GDELT), Reuters Corpora, 20 Newsgroups and Fast.ai are vastly used for categorizing text.

5. Evaluation metrics:

The motivation behind recommender system intends for predicting the chances that the user enjoys an unknown entity presented to them. As this task depends on whether new items recommended are according to preferred patterns or not, the main focus remains on studying and knowing the preferences patterns of every user. Traditionally similarity based metrics were used to determine the 'nearness' of new items. Business related metrics are click through rates, adoption measure (suggestions effectiveness), sales, revenue, distribution patterns and user engagement. Here based on recommendations tasks, we classify the evaluation metrics (figure-2).

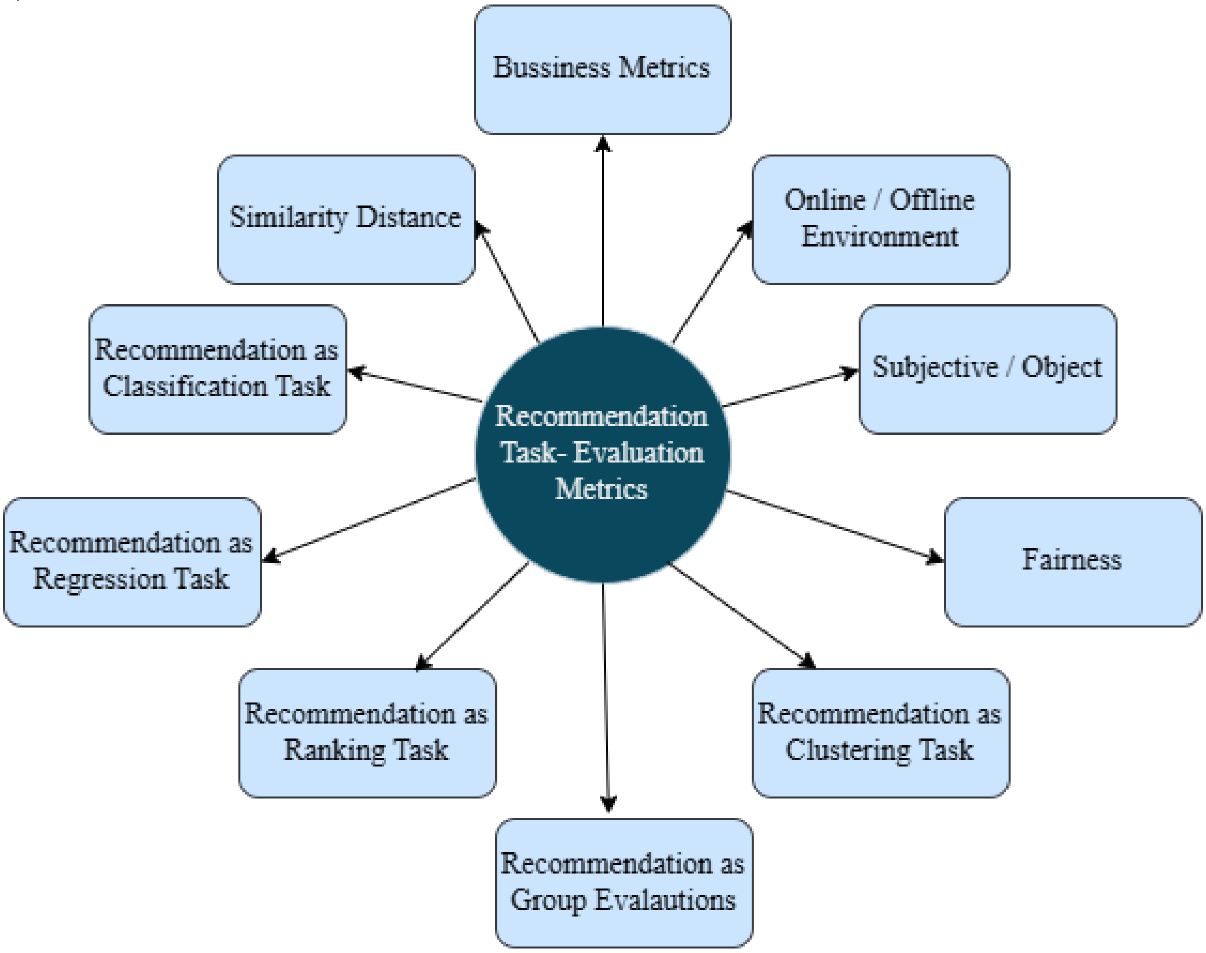

Figure-2 Categorization of evaluation metrics based on tasks

5.1 Recommendation as classification task:

The tasks that view recommendation as classification problem generally use metric like accuracy, precision, F1 score, recall and Area Under Curve (AUC) score. AUC or Receiver operating characteristic (ROC) score denotes the probability of higher ranking of positive entities i.e., probability curve representing sensitivity vs 1-specificity. Sensitivity or recall or true positive rate is the fraction of positives classified correctly. Specificity is the true negative rate stating fraction of negatives classified correctly. 1-specificity is the false positive rate stating negative class proportions classified correctly. Thus AUC curve with high value of x-axis show false positives (FP) are greater in number than true negatives (TN), and high value of y-axis show true positives (TP) are greater in number than false negatives (FN). Precision states fraction of true samples classified correctly with true samples. MCC (Matthews Correlation Coefficient) correlates actual true classes and predicted labels. Instead of quantifying the difference between predicted and expected probability distribution as in log loss, Brier score (BS) quantifies the average difference in just probabilities of those, i.e., accuracy of probabilities along with MSE.

$$\text{Sensitivity}=\text{Recall}=\text{TPR}=\text{TP} / \text{TP}+\text{FN} \quad (1)$$

$$\text{Specificity}=\text{TNR}=\text{TN} / \text{TN}+\text{FP} \quad (2)$$

$$1\text{-specificity}=\text{FPR}=\text{FP}/ \text{TN}+\text{FP} \quad (3)$$

$$\text{Accuracy}=\text{TP}+\text{TN} / \text{TP}+\text{TN}+\text{FP}+\text{FN} \quad (4)$$

$$\text{Precision}=\text{TP} / \text{TP}+\text{FP} \quad (5)$$

$$\text{F1} = (2*\text{precision}*\text{recall}) / (\text{precision}+\text{recall}) \quad (6)$$

$$\text{Gini index}= \sum_{i=1}^{n}(2i-n-1)x_i / n\sum_{i=1}^{n} xi \quad (7)$$

Where x denotes values observed, n is total of those values, and i is the rank of values.
[15] calculated harmonic mean to know the performance of both users and creators (C&C):

$$C\&C= (2 \times (1 - Gini) \times F1)/ ((1 - Gini) + F1) \quad (8)$$

$$MCC\ Score = TP \times TN - FP \times FN / \sqrt{((TP + FP)(TP + FN)(TP + FP)(TN + FN))} \quad (9)$$

$$BS= 1/ N \sum_{i=1}^{N} (pf_i-po_i)^2 \quad (10)$$

Where N denotes total instances present, po_i is actual expected probability of given instance i and pf_i is predicted probability.

5.2 Recommendation as regression task:

The tasks that view recommendation as regression problem use metrics like MAE, MSE, RMSE, PCC, RMSLE, R2, adjusted R2. Mean absolute error (MAE) is difference in actual and estimated values. Mean squared error (MSE) is square of MAE. Root mean squared error (RMSE) is root of MAE. Root mean squared log error (RMSLE) is log of RMSE. R2 or Coefficient of Determination or Goodness of fit is context independent metric to evaluate goodness of regression line as compared to mean line. Pearson Correlation Coefficient determines the relationship strength between variables.

$$MAE = 1/n \sum |y - y'| \quad (11)$$

$$MSE= 1/n \sum |y - y'|^2 \quad (12)$$

$$RMSE= \sqrt{[1/n \sum_{k=1}^{n}(yi - y'i)]} \quad (13)$$

$$R2= 1- (SS_r/SS_m) \quad (14)$$

$$Adjusted\ R2= 1- [(n-1 / n-k-1) * (1-R^2)] \quad (15)$$

$$PCC= (E[XY] – E[X]E[Y]) / (\sqrt{(E[X^2]-E[X]^2)}\ \sqrt{(E[Y^2]-E[Y]^2)}) \quad (16)$$

Where y is actual output and y' is predicted output. SS_r is regression line's squared sum error and SS_m is mean line's squared sum error. R2 is worst in case when both lines overlap. For PCC, the numerator term is covariance and denominator is the standard deviation of variables X and Y.

5.3 Recommendation as ranking task:

For this ranking task, commonly used metrics are Hit ratio (HR), Average precision (AP), Mean Reciprocal Rank (MRR), Discounted Cummulative Gain normalized (DCG), normalized Discounted Cummulative Gain (nDCG). MRR or average reciprocal hit ratio (ARHR) is obtained from weighting through reciprocal rank of relevant scores of top n entities and then summing up them

$$AP= \sum_{n} (R_n – R_{n-1}) P_n \quad (17)$$

$$HR= \text{total users for whom top n recommendation is correct / total users in test set} \quad (18)$$

$$RR(k)= \sum_{i \le n} relevance_i / ranking_i \quad (19)$$

$$MRR= 1/k_{all} \sum_{k=1}^{k(all)} RR(k) \quad (20)$$

Where P is precicion and R is recall

DCG shows level of information gain on prediction set. Next by sorting documents according to relevance rel_i based on position p. But as performance cannot be compared based on a single query or critera, nDCG scales the results considering the best among analysed.

$$DCG= \sum_{i=1}^{n}(ri) / \ln(i + 1) \quad (21)$$

$$IDCG= \sum_{i=1}^{rel(i)}(ri) / \ln(i + 1) \quad (22)$$

$$nDCG= DCG / IDCG \quad (23)$$

5.4 Recommendation as Clustering task and Group evaluations:

While considering clustering, the way to find out the optimal number of clusters or to select the optimal value for the number of clusters to be analysed, techniques like Average Silhoute Criterion (a measure of similarity of an entity to its cluster compared to other ones), Calinski Harabasz [84] (sum of inert clusters and intra cluster dispersions) and Elbow method (plot observed variation as function of k number of clusters and select elbow curve for use) are applied.

V score measures the meaningfulness of existing clusters. [86] First, Homogeneity h gives the scenario of which and upto what extent are the samples in clusters are similar, using Shannon's entropy. Next completeness c shows which of similar samples are accurately clustered. For the same, H(X/L) is found in same way as H(L/X). Then to decide whether algorithm is good or not, Normalized Mutual Information (NMI) is calculated between homogeneity and completeness for V measure.

$$H(L/X) = - \sum_{l,x}^{0} [(n_{lx} / N) \log n_{lx} / n_x] \quad (24)$$

$$h = 1 - H(L/X) / H(L) \quad (25)$$

$$c = 1 - H(X/L) / H(X) \quad (26)$$

$$NMI = 2 * [(h*c)/(h+c)] \quad (27)$$

Where, n_{lx} / n_x, depicts ratio of total samples in cluster x labelled as l and total existing samples of cluster k

$$\text{Silhouette Score} = (v-u) / \max(u,v) \quad (28)$$

$$\text{Inter cluster dispersion } A = \sum_{i=1}^{n} m_i * \|C_i - C\| \quad (29)$$

$$\text{Intra-cluster dispersion } B = \sum_{i=1}^{n} \sum_{j=1}^{mi} \|M_{ij} - C_i\|^2 \quad (30)$$

$$\text{Calinski Harabasz (CH) index} = [A/(n-1)] / [B/(X-n)] \quad (31)$$

Where m_i is total observations in cluster i, X is total observations, M_{ij} is jth observation in cluster i, n is total number of clusters, C_i is centroid of cluster i and C is Centroid of data. For Silhouette Score, u is the average of intra-cluster distance (average of distance among each point present within given cluster.) and v is average inter-cluster distance (average of distance between total present clusters.)

Apart where, group score s are required, various measures like Group prediction score [25], Group divergence, consensus function and T testing have been applied. Group prediction score is obtained by combining predictions at individual levels present in specific group. Group Divergence (dis) shows degree of dissimilarity in predictions at individual levels in a group.

$$GP(G, i) = 1 / |G| \left(\sum_{u \in G} P(u,i) \right) \quad (32)$$

$$dis(G, i) = 1 / |G| \left(\sum_{u \in G} P(u,i) - mean(G,i) \right)^2 \quad (33)$$

$$\text{Consensus function: } F(G, i) = w1 \times GP(G, i) + w2 \times (1 - di\ s(G, i)). \quad (34)$$

where mean(G,i) denotes average of predictions at individual level in a group G for document i, w1 and w2 are group wise predictions and disagreement. Weights w1 and w2 sum up to 1.

P value tells about chances of getting outcomes when null hypothesis is true. Although p values doesn't hold any significance or numeric score individually, it tells a lot about the alignment of a score to a specific sample. To know about any notable distinctness between means of groups, T test is used [87].

$$t = \text{variance of mean between groups} / \text{variance of mean within groups} \quad (35)$$

5.5 Fairness in recommendations:

The term fairness refers to unbiased recommendations with respect to either person or group of customers or creators or providers. [75] Mentioned two type of fairness: Process and Outcome, further categorizing outcome fairness to eight subclasses: individual, group, consistent, calibrated, counterfactual, envy free, Rawlsian maximum and maximum shared. Individual and group are based on target whereas rest all are based on concept. For all these fairness, different metrics are discussed including Kolmogorov-Smirnov statistic, rKL, rRD, rND, Pairwise Ranking Accuracy Gap (PRAG), variance, Min-Max Difference (MMD), F- statistic (ANOVA), Gini coeicient, Jain's index, Min-Max Ratio, entropy, Least Misery, KL-divergence, MinSkew, MaxSkew, JS-divergence, Overall Disparity, L1 norm and envy freeness.

Besides these, Equity Attention for group fairness (EAGF) measurement and a Supplier Popularity Deviation (SPD) are used.

$$EAGF = \sum_{k=1}^{|g|} \sqrt{(|L(k)|)} \quad (36)$$

$$SPD = \sum_{k=1}^{|g|} \left(\frac{|L(k)|}{|L|} - \frac{|A(k)|}{|A|} \right) / |g| \quad (37)$$

Where for EAGF, *g* represents the author group, (K) denote recommendation set of items for any group k. For SPD, *L* is recommended item set, (K) is item set from training data of group k and *A* is training set.

5.6 Subjective / Objective recommendations:

Subjective measures tells about recognizing and assessing the consciousness of user emotions, whereas objective measures give information on determining both conscious and unconscious emotions. Being more precise, physiological or unconscious responses such as liking new items or tending towards unexplored entities or general preference [29], falls under objective measures. While for subjective measures, generally conscious emotions of users like users' trust [29], their self-study and self-generated reports as well as timely question answers. Objective measures include accuracy metric along with diverse and assorted estimations, user exposure and coverage to items, demographics with their interrelations, as well as novelty with relevance to users. On the other side, subjective measures are judged by humans based on ratings or questionnaires from users' side while for creators these measures include clicks, dwell time, shares, dislikes etc.

Metrics like Intra List Similarity (ILS), user coverage (UC) [83], diversity [83] and novelty [83]. ILS gives average of similarity in pairs where similarity is found based on particular's metadata and context. Diversity is given by subtracting similarity for user set from one. For novelty, first overall rankings of domain are considered and then from each user's top n prediction lists average of all popularity rankings for that individual's prediction is taken.

$$\text{UC}= [\sum_{u=1}^{n}(\text{Hits}(\text{predicted rating} \geq \text{threshold}))_u] \text{ / number of users} \quad (38)$$

$$\text{ILS for user}_x= 1/n\ [\sum_{u=1}^{n}\sum_{v=1}^{n} \text{sim}(i_u,i_v)] \quad (39)$$

Where, n is top n predictions for each user u. For ILS, n is top n recommendation for two different items u and v. $\text{sim}(i_u,i_v)$ is similarity between u and v.

5.7 Recommendation in Online / Offline Environment:

For online evaluations, considering required relevant parameters, some specific like Click through rate (CTR) saying about number of users that clicked link and those who are viewing after clicking can reveal online scenario. Apart from it, dwell time tells about time spent by user on that link. Bounce rate says about returning time to main engine or closing time of particular link. Furthermore, average time duration spent on link, session durations and repetition in sessions are also accordingly considered for online evaluations of these systems.

$$\text{CTR= total clicks / total impressions} \quad (40)$$

Focusing the offline evaluations, measures like DCG, nCDG, MRR, MAP are widely applied. Mentioned measures for offline evaluations generally reveal the top k entities for recommendation from the available set considering the history of user behaviour but cannot reveal information about new set or some another possible set of recommendations that user will definitely like.

Here is the summary of detailed analysis of evaluation metrics with their specific target task showing the main aspects or ideas to be taken care of while using that particular metric. Given table 8 shows summary of characteristics of evaluation metrics.

Table-8 comparison of goals and characteristics of evaluation metric based on tasks done

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	Ranking

	parameters and recommending same number of items), checks whether independent variable contributes to model or not.	
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

6. Analysis:

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.

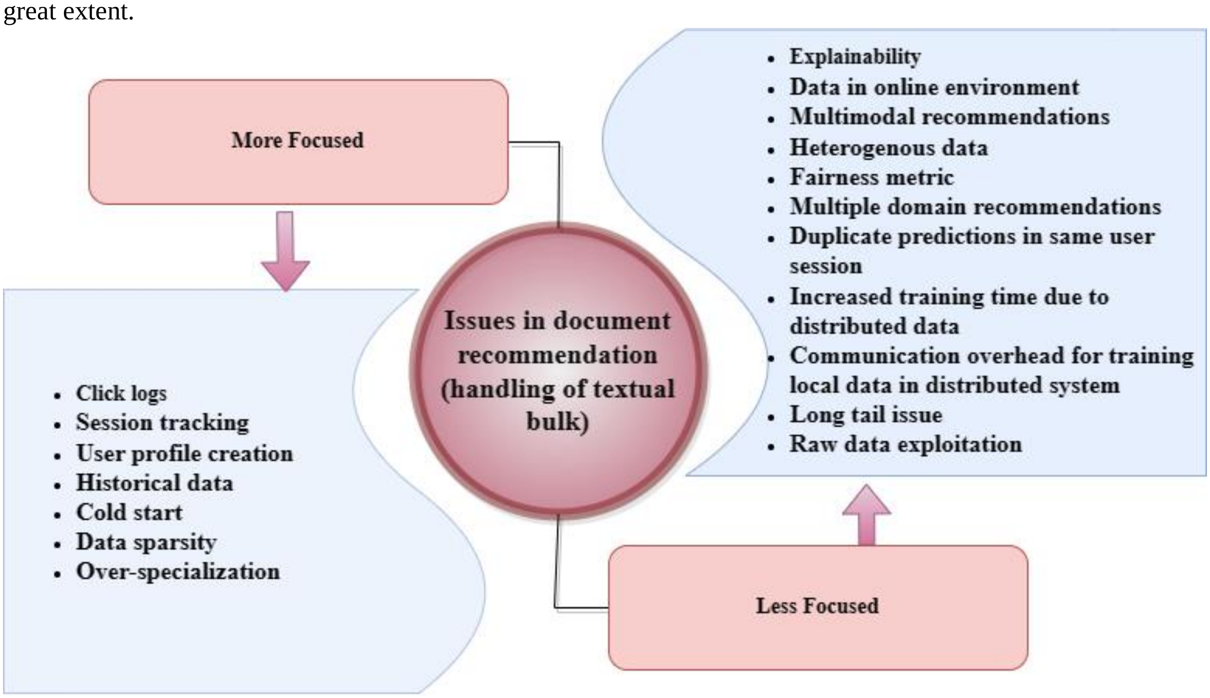

Figure-3 Issues or points in document recommendation (textual bulk handling) which are commonly explored as well as less explored and needs attention.

7. Challenges:

Security and privacy: Data for recommendation are prone to various attacks like online adversarial and data poisoning which can inject crafty fake information or delete some important parameters which can affect model parameters and training with undesirable estimations. This in turn can create biases towards entities, manipulating users' exposure and fakery domination over the market. [31] and [89] gives concept of storing public data on central server and users' data locally on their side. They locally compute the model by exchanging required data and then updating server at the end. During exchange of required data, to prevent leakage, instead of directly sending user embeddings, decomposed embedding linearly combined basis for each user embedding where weights of combinations are secured by Differential privacy locally. This solves issue to some extent as only gradients are sent to update server, still this induces communication overhead and are however vulnerable to model attacks. Besides profiling phase is the high time when information of users are at peak exposure. Although protocol like General Data Protection Regulation is present but it cannot tell about misuse of data by public organizations which show proper need of regulations and rules for privacy.

Anonymization of data: Anonymised data can solve privacy issues. But along with it, steps in some problems like reduction in accuracy. To start with, for this purpose access to cookies [91] is required but if the user switches to incognito mode. This technique is tough to implement. . [93] Explored use of anonymised data with homophormic encryptions for privacy and elimination of third party dependency and participation. Considering the risk, some anonymisation techniques can be applied but are unsuitable for high dimensional data [92], where sensitiveness of each attribute has to be considered. Along with using anonymisation, proper data masking and categorization of sensitive attributes is must to ensure good performance [91]. In addition, it is quite difficult to recognize data patterns from anonymised data and for recommender systems proper data insights is the lifeline. Furthermore, anonymisation don't provides reindentification of encrypted entities, whereas Pseudonymization enables the same.

Bias and Filter bubbles: Customer behaviour and their consumption information is most important aspect of business analytics [85]. Generally recommendations are done on basis of past historical behaviour and deductions based on the same. This observation bias can lead to filter bubbles, which means users are introduced to limited content which supports and nourish their mind set and beliefs. This can in turn lead to supporting extremism of self-thoughts, creating echo chambers disregarding alternative ideas. In addition, sometimes imbalanced data lead to biasness where cricket, kabaddi, or sports recommendations are common for males and festival discount offers-sales are generally recommended to females. Although it's necessary to analyse consumption behaviour but its overspecialization induces filter bubbles.

Fairness: General analysis show system performance and accuracy have conflict with the fairness [75]. But for optimal solutions it is necessary to have all these. Many of works like [90] explores fairness, but of only single type. Adaption of joint, all-purpose, and overall fairness with detailed analysis of effects of various evaluation metrics still is a topic to explore. For correct tackling of fairness, identification of sensitive attributes and its proper usage holds equal importance. Existing works are carried out based on assumption that required attributes of fairness are present in data, whereas it may happen related required information is missing which shows necessities of fairness related investigations. Explainability for fairness in recommendation task can convince users in better way increasing their level of satisfaction and consecutively ensuring the trust for system [75]. [32] Deployed adversarial learning to recognize bias free (learn bias independent attributes) and aware (sensitive attributes) embeddings of users, thus mitigating unfairness due to bias. But their hyper parameter tuning to remove bias and increase fairness is quite difficult which can expose the model for penetration. Fairness doesn't imply equality in content to all users, frequently shopping people are highly exposed to good estimations as compared to people who rarely shop.

Serendipity: This concept is amalgamation of characteristics like pertinence of entity to user (degree of connectedness or degree of usefulness to a user), entities out of blues or surprising ones (unexpected ones but seeming fascinating, attentiveness) and novelty (item that has been forgotten by user or is not exposed to yet but tend to it once knowing about it) [29]. As an example it can be said that any technical content is novel if a user hasn't seen it before and reads it but it isn't of serendipity if he doesn't wants to work on it. In that case, serendipity is when that particular user is interested in further working on that topic after reading it [29].

Time Factor: The recommender systems must respond to behaviours of user in specific time period. This states the need of real time processing, low latency, efficient and faster computations of bulky data. Along with this, recommendations on proper time i.e., in context of trends, popularity, seasonal behaviours, idiosyncratic demands, recency and out of ordinary estimations matter a lot. For example, news tends to be outdated faster than technical documents or technical content. As well as the consumption time or time spent too are different for both of them. News clicks generally have less dwell time whereas time spent on technical content is quite more.

Duplications: This occurs when many of the links contain some content (either leading to the same document or page or containing the same content semantically). This can lead to users losing interest and trust on the system. Or sometimes it may happen that link actually contains new, worthy information for users but they may misconceive it as trashy. This can trap users to limited content minimizing their exposure. For this problem, explainability of recommendations can be an effective solution.

Diversity: For enhancement of user experience, trust, interest and involvement, it is most important to present them with variety and multifarious information. As mentioned earlier, users tend to lose interest when presented with links or content having homogeneous information. Commonly many works focus on ways to optimize the recommendations task with accuracy metric by brushing aside the matter of diverse content [43]. Diverse recommendations not only mean that new content estimations are non-identical to the previous recommendations, but also expose users to diverse heterogeneous top n interests accordingly. As diversity evaluates distinctness and variance among predicted items, it is generally implemented with help of re-ranking methods like Intra list similarity, normalized diversity or temporal diversity [29]. Although diversity is used with re-ranking concept, there exists trade-off between re-ranking's accuracy due to high computations and diversity. Possible directions for advancement in diversity works include topic of trade-off for accuracy-diversity, analysing level of diversity according to user [29] and making system scalable for diversification.

Incognito tabs: Usage of incognito tabs by users becomes a barrier in the retrieval of some crucial user information through cookies, history tracking, and logs. As users' interests aren't updated frequently, it can result in homogeneous estimations consequently losing users. Therefore before using recommendation system for predictions, further work includes a mechanism to identify and handle or retrieve user data in case incognito tabs are used.

Content Validity and Control: Spam or fake content many a times misguide user. All the data that is being recommended must be up to date, validated, and from authorized sources. Before recommendations low quality or harmful content must be removed. The need of tackling the misinformation or content that lead to negative impact socially is required.

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.

Multilingual and Multi-Modal: Document recommendation usually can include images or links to some other content that could be in any format. Apart it can be in different languages too [29]. Thus current research lacks mechanism to handle multi lingual data. Most of the work deal with single kind of data like either text or image or video or links, but have ignored the hurdles caused when there is an involvement of two or more types. For example: for technical document recommendation, features can be in any form apart from text like image (in form of work flows or equation), or links (leading to codes, video etc).

Other issues: Sequential consumption of entities can be useful in case where a user wants to get updated for current information or wants some more inline information to current reading if published in future. Sequential consumption wouldn't be helpful if it recommends similar items frequently like for music or movies [29]. The unavailability of a benchmark dataset that can measure all the needed measures or metrics including "total fairness" for every concerned partaker of the system is still an issue. Besides, churn rate of users too gives much information on their satisfaction levels, rate at which recommendation changes once user gives ratings to new entities, parameters affecting them and interests. Using this information, churn rate can be minimized increasing user trust. Apart from general metric accuracy, it is necessary to efficiently deal with missing data or noise. System's capability in making relevant estimations though noise and outliers are present determines the robustness of the system. Finally, proper handling of the issue of data sparsity and cold start when any new entity either user or item enters is still pending to be explored in online environment.

8. Discussions and Future Scope:

Deep Reinforcement Learning: Use of deep reinforcement learning provides insights to internal working of models enabling analysers to know about information like consumer behaviours, explanations on decisions made. This feature can assist in monitoring the training phase so that if anything undesirable occurs, it can be handled with time. Use of these methods can reveal recall of model features at any given stage. As this provides a deep analysis and access of internal states, these methods are inappropriate where security and privacy are viewed as a primary issue. These methods are most suitable for unstructured environments. Deployment of SHAP (SHapley Additive exPlanations) values permits to analyse the long term effects of state vs action combinations for individual user. RL-SHAP plotting can show representation stating positive, negative, high or low effect of features on the action.

Explainability and Interpretability: When security is given more importance, obviously underlying ML/DL models has to be non-transparent. Due to this non-transparency, it lacks proper required comprehension of information resulting in challenge for ML/DL developers to pinpoint failures and appropriate reasons for them. Scope in this case includes usage of different frameworks [95] like SHAP, LIME (Local Interpretable Model-agnostic Explanations), Shapash, InterpretML, ELI5 and OmniXAI (Omni eXplainable AI).

Blockchain: A gap has been created between the model performance and it security. Promoting feature like resiliency, durability, flexibility and trust, blockchain can contribute a lot to privacy-preserving recommender systems eliminating third party interference. The solutions to commonly looked-for issues like transparency and decentralized training include the possible directions of work along with providing security. But things to be taken care while using blockchain include scalability, distributed training and explainability [96].

Multi task and transfer learning: As recommendation tasks contain two or more objectives, for which in turn arises requirement of two or more losses for its optimization. MLT provides this optimization permitting the sharing of variables among task, enabling transfer learning. [29] Pointed further use of TL for data sparsity. But due to difference in features of datasets used, tackling mechanism for noise, outlier and missing information needs to be considered.

Federated Learning: There is always a restriction of underlying hardware when it comes to online computations, arising need for decentralized management. In addition, on device computations may have limitations in data, its labelling and its model size. Dealing with heterogeneous dives, edge computations and their optimization, data sensibility issue has been tactfully handled using FL. If heterogeneity of system, heterogeneity of statistics, privacy concerns and communication overhead considered properly [97], FL has much potentials and future scope in online environment.

Augmentation and Contrastive learning: To start with, image features for documents are generally ignored, further its augmentation related to required environment is studied seldom. As it is self- supervised, it can also work efficiently in online environment. Due to its nature of learning without any labels, CL has unexplored scope for this type of recommendation task.

Raw data exploitation: As self-supervised methods are prevailing, exploration and analysis of directly raw data can serve with more supervising prompts for relevant model training. This task starts with the proper studying of augmentation for the pre-training purpose. Proper exploitation of augmentation can encourage task of learning without labels. Following augmentation comes the main job of making recommendations, but again this pre-trained model with augmentation can be utilized to enhance recommendation performance further.

Multi-objective reinforcement learning (MORL): MORL are for multiple objectives. Although RL is a strong exemplar for handling uncertainties in sequential estimations, RL tend to maximize just some selected numerical values portraying single objective which is long term [99]. Even if multiple objective are taken onto account, there are viewed as linear combination. Factors need to keep in mind while conceptualizing solutions are single/ multiple policies, utility functions nature, deterministic/ stochastic policies, scalarised expected returns (SER) and expected scalarised returns (ESR) according to [99]. This serving as powerful technique can make systems adaptable for real time.

Causal Bayesian networks (CBN): Commonly used metrics for fairness considers relation between sensitive features and output. But in contrast, training data can have patterns leading to unfairness that could have effect on other attributes too. Causal graphs or CBN can assist in interpreting the process and reason of sensitive attributes effects on others. But first problem is of construction of casual graphs. General information used for its design is ID details [75]. Considering more features leads to complications for construction. Adding to these, another challenging aspects include taking care of extraneous determinant and dealing with complexities of performing counterfactual inference [98]. Yet there is much opportunity to explore these drawbacks.

9. **Conclusion:**

This study concerns on different categorizations of document recommendation task along with their features considered. Here we have explored the most recent researches and literature works for specially document recommender system as well as have analysed and compared different techniques, models according to categorized task along with principle feature or attributes contributing for that specific task. Apart from these, we show comparisons of various target or goal of literatures for that particular category. In addition to that, we comprehensively present the datasets and evaluation metrics. Stating analysis of evaluation metrics, we observed some key points to make recommendation system more responsible. Figuring out the challenges and showing possible unrealized potential for future work, we unveil the most recent and up to date prospective and insights for extending research in this field. The knowledge and understanding provided through this work can aspirate researchers for progressive and state of art solutions for current domain as well as other real time application domains.

List of Abbreviations:

P-LDRS: Pre-Learned Legal Document Recommender System
GRU: Gated Recurrent Unit
RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers
LDA: Latent Dirichlet Allocation
SCA: Sentence Complexity Analysis
CAOT: Citation Analysis Over Time
KDM: Keyword Diversity Measurement
SQM: Scientific Quality Measurement
SEAN: Social Explorative Attention Network
SMOTE: Synthetic Minority Oversampling Technique
ACM: Association for Computing Machinery
IPTC: International Press Telecommunications Council
GDELT International Press Telecommunications Council
MIND: Microsoft news dataset
CTR: Click Through Rate
ORP : Open Recommendation Platform
CF: collaborative filtering
CBF content-based filtering
MF: matrix factorization
NMF: Non-negative matrix factorization
PMF: Probabilistic matrix factorization
BPR: Bayesian personalized ranking
GLM: Generalized linear modeling
RARD: Related-Article Recommendation Dataset
lg2vec: log2vec
AUC: Area Under Curve
ROC: Receiver operating characteristic
MCC: Matthews Correlation Coefficient
BS: Brier score
MAE: Mean absolute error
MSE: Mean squared error
RMSE: Root mean squared error
PCC: Pearson Correlation Coefficient
HR: Hit ratio
AP: Average precision
MRR: Mean Reciprocal Rank
DCG: Discounted Cummulative Gain
nDCG: normalized Discounted Cummulative Gain
ARHR: Average Reciprocal Hit Ratio
IDCG: Ideal Discounted Cummulative Gain
CH: Calinski Harabasz index
PRAG: Pairwise Ranking Accuracy Gap
MMD: Min-Max Difference
ANOVA: Analysis of Variance
rND: Normalized discounted difference
rKL: Normalized discounted KL-divergence
EAGF: Equity Attention for group fairness
SPD: Supplier Popularity Deviation
ILS: Intra List Similarity
UC: User Coverage
MORL: Multi-objective reinforcement learning
CBN: Causal Bayesian networks

Declarations:

Ethical Approval:

- Not Applicable

Funding' and/or 'Competing interests:

- The authors did not receive support from any organization for the submitted work. (As this work is part of full time PhD)
- The authors have no relevant financial or non-financial interests to disclose as well as no competing interests to declare that are relevant to the content of this article.

Authors' Contributions:

- This manuscript, figures and tables are prepared by Aanal Raval. Dr. Komal Borisagar has guided during the process.

Conflicts:

- All authors have no conflicts

Availability of data and materials:

- Not applicable

References:

[1] Jenish Dhanani, Rupa Mehta, Dipti Rana, "Effective and scalable legal judgment recommendation using pre-learned word embedding", Complex & Intelligent Systems, Springer, 28 January 2022, https://doi.org/10.1007/s40747-022-00673-1

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

[3] Nithirun Numnonda, Sapa Chanyachatchawan & Nuengwong Tuaycharoen, "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, "Research on application of article recommendation algorithm based on Word2Vec and Tfidf", 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)

[5] Stefano Ferilli , and Liza Loop, "Toward Reasoning-based Recommendation of Library Items – A Case Study on the e-Learning Domain", 18th Italian Research Conference on Digital Libraries (IRCDL 2022), February 24–25, 2022, Padua, Italy, CEUR Workshop Proceedings.

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

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

[8] Duc H. Le, Tram T. Doan, Son T. Huynh, and Binh T. Nguyen, "SimCPSR: Simple Contrastive Learning for Paper Submission Recommendation System", arXiv:2205.05940v1 [cs.IR] 12 May 2022

[9] Aram Karimi, Simon Dobnik, "Multi-task recommendation system for scientific papers with high-way networks", arXiv:2204.09930v1 [cs.IR] 21 Apr 2022

[10] Nianlong Gu, Yingqiang Gao & Richard H. R. Hahnloser, "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, "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, Grinrey, 2021

[12] Vilmos Tibor Salamon, Tsegaye Misikir Tashu & Tomáš Horváth , "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. "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, "Context-Based Quotation Recommendation", Association for the Advancement of Artificial Intelligence, 2021, Google Research

[15] Wenyi Xiao, Huan Zhao, Haojie Pan, Yangqiu Song, Vincent W. Zheng, Qiang Yang, "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. "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, "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, "Recommendation Systems Based on Textual Document Analysis", 2021 29th National Conference with International Participation (TELECOM), IEEE

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

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

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

[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, "Content Recommendation Based on Topic Modeling", Computational Methods and Data Engineering, Advances in Intelligent Systems and Computing 1227, Springer Nature Singapore Pte Ltd. 2021

[24] Raghvendra Rao, Suyash Choubey, Georg Gutjahr, and Prema Nedungadi, "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, "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, 2019.

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

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

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

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

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

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

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

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

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

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

[37] Tao Qi, Fangzhao Wu, Chuhan Wu, Yongfeng Huang, "Personalized News Recommendation with Knowledge- aware Interactive Matching", In

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

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

[40] Mingyuan Ma, Sen Na, Hongyu Wang, Congzhou Chen, Jin Xu, "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, 2021.

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

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

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

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

[46] Jiasen Lu, Dhruv Batra, Devi Parikh, Stefan Lee., "Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks.", 2019

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

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

[49] Jon Atle Gulla, Lemei Zhang, Peng Liu, Özlem Özgöbek, Xiaomeng Su "The Adressa Dataset for News Recommendation", Dep. of Computer Science, NTNU, ACM ISBN 978-1-4503-4951-2/17/08

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

[51] Hacker News, https://www.kaggle.com/datasets/hacker-news/hacker-news, last accessed 2023/04/10

[52] Language Data, https://webscope.sandbox.yahoo.com/catalog.php?datatype=l, last accessed 2023/04/10

[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. MIND: A Large-scale Dataset for News Recommendation. ACL 2020.

[54] MIND: Microsoft News Dataset, https://msnews.github.io/, last accessed 2023/04/10

[55] Fake News Dataset, https://www.kaggle.com/datasets/sumanthvrao/fakenewsdataset, last accessed 2023/04/10

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

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

[58] Palanigurupackiam Nagaraj, Perumalsamy Deepalakshmi, "An intelligent fuzzy inference rule-based expert recommendation system for predictive diabetes diagnosis", Internatioanl Journal of Imaging systems and Technology, Wiley 03 February 2022.

[59] Ghosh, Chinmoy and Saha, Subrata and Saha, Soumyadeep and Ghosh, Niladitya and Singha, Kumaresh and Banerjee, Annwesha and Majumder, Sourav, Machine Learning Based Supplementary Prediction System Using K Nearest Neighbour Algorithm (January 10, 2020). 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, July 2021, 113546

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

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

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

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

[66] Danyang Liu, Jianxun Lian, Zheng Liu, Xiting Wang, Guangzhong Sun, Xing Xie, "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, "A hierarchical recommendation system for E-commerce using online user reviews", Electronic Commerce Research and Applications, Elsevier, Volume 52, March–April 2022, 101131

[68] JIZHOU HUANG, HAIFENG WANG and WEI ZHANG, TING LIU, "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, July, 2020

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

[70] XinMei, XiaoyanCai, SenXu, WenjieLi, ShiruiPan, LibinYang, "Mutually reinforced network embedding: An integrated approach to research paper recommendation", Expert Systems with Applications, Volume 204, Elsevier, 15 October 2022, 117616

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

[72] Nicola Capuano, Luigi Lomasto, Andrea Pozzi & Daniele Toti, "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, 2021

[74] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Jundong Li, Zi Huang, "Self-Supervised Learning for Recommender Systems: A Survey", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2022

[75] YIFAN WANG, WEIZHI MA, MIN ZHANG, YIQUN LIU, SHAOPING MA, "A Survey on the Fairness of Recommender Systems", ACM Trans. Inf. Syst, 2022

[76] RefSeer data, https://github.com/chbrown/refseer, last accessed 2023/04/10

[77] arXiv Dataset, https://www.kaggle.com/datasets/Cornell-University/arxiv, last accessed 2023/04/10

[78] JRC-Acquis, http://data.lynx-project.eu/it/dataset/jrcacquis, last accessed 2023/04/10

[79] AMiner (database), https://en.wikipedia.org/wiki/AMiner_(database), last accessed 2023/04/10

[80] DBLP Dataset, https://paperswithcode.com/dataset/dblp#:~:text=The%20DBLP%20is%20a%20citation,year%2C%20venue%2C%20and%20title., last accessed 2023/04/10

[81] Skill2vec Dataset, https://paperswithcode.com/dataset/skill2vec, last accessed 2023/04/10

[82] RARD: The Related-Article Recommendation Dataset [Data], https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/HA8EAH, last accessed 2023/04/10

[83] Recommendation System Evaluation Metrics, https://medium.com/@rishabhbhatia315/recommendation-system-evaluation-metrics-3f6739288870, last accessed 2023/04/10

[84] Calinski-Harabasz Index for K-Means Clustering Evaluation using Python, https://towardsdatascience.com/calinski-harabasz-index-for-k-means-clustering-evaluation-using-python-4fefeeb2988e, last accessed 2023/04/10

[85] Raval, A.S., Oza, R.K. and Jain, A.V., (2021) Social Media Content Classification and Detection for Trend Analysis Using Machine Learning Techniques., IJSRSET

[86] V measure: an homogeneous and complete clustering, https://towardsdatascience.com/v-measure-an-homogeneous-and-complete-clustering-ab5b1823d0ad, last accessed 2023/04/10

[87] T-test, https://www.geeksforgeeks.org/t-test/, last accessed 2023/04/10

[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, “Towards a more reliable privacy-preserving recommender system”, Information Sciences, Elsevier, Volume 482, May 2019, 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] Anonymisation, https://www.imperva.com/learn/data-security/anonymization/, last accessed 2023/04/10

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

[95] Top Explainable AI (XAI) Python Frameworks, https://moez-62905.medium.com/top-explainable-ai-xai-python-frameworks-in-2022-94ff4610b0f5, last accessed 2023/04/10

[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] Introduction to Federated Learning and Challenges, https://towardsdatascience.com/introduction-to-federated-learning-and-challenges-ea7e02f260ca, last accessed 2023/04/10

[98] Causal Bayesian Networks: A flexible tool to enable fairer machine learning, https://www.deepmind.com/blog/causal-bayesian-networks-a-flexible-tool-to-enable-fairer-machine-learning, last accessed 2023/04/10

[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