Babylonian Journal of Internet of Things
Vol.2024, **pp**. 161–168
DOI: https://doi.org/10.58496/BJIoT/2024/018; ISSN: 3006-1083
https://mesopotamian.press/journals/index.php/BJIoT

An Author-Centric Framework Error Minimization in Scholarly Recommender System (Acfemsr)

A.H. Zaharaddeen[1] *, , Nurra Mukhtar [1],

[1]*Department of computer science, Faculty of Natural and Applied Science, Umaru Musa Yaradua University Katsina (UMYU), Nigeria.*

ARTICLE INFO

Article History

Received 07 Aug 2024
Revised: 06 Oct 2024
Accepted 08 Nov 2024
Published 05 Dec 2024

Keywords

Recommendation Error

Mean Absolute Error (MAE)

Root Mean Square Error (RMSE)

Scholarly Recommender Systems

Error Minimization

ABSTRACT

In the current landscape of computational innovations, the internet hosts a vast repository of publications sourced from various channels. Consequently, researchers encounter unprecedented challenges in identifying publications relevant to their research interests. Navigating through the multitude of options provided by search engines is not only impractical but also prone to selection errors – termed ***Recommendation Error (RE)***. These experiences underscore the need for novel research avenues. Previous studies have proposed numerous recommendation frameworks exclusively for scientific paper recommendations. However, many of these approaches have been plagued by RE, thereby compromising the integrity of recommendation systems. Recommendation Errors typically arise from underutilization of key features in the recommendation process. To mitigate Recommendation Errors, this study leverages publicly available metadata features, including Title, Abstract, Author(s), and keywords. Feature vectors for each candidate paper (CP) and the paper of interest (POI) are computed using CountVectorizer, and cosine similarity formula is employed to identify similar papers suitable for recommendations. The effectiveness of the proposed framework is evaluated using Error evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Finally, the proposed Framework is compared against two previous Baseline approaches. The evaluation results demonstrate that the proposed approach exhibits lower Recommendation Errors compared to its baseline counterparts. Additionally, this research highlights the Author(s) feature as the most influential among the four features utilized by the Proposed Framework.

1. INTRODUCTION

The integrity of any recommender system lies over the accuracy of its recommendation and its ability to present less /error-free recommendation list to the user. Current collaborative-based recommendation systems are simply erroneous due to the cold start problem [1]. Again, recommendation RE occurred due the uncertainty of data distribution especially in collaborative system that relied heavily on user's ratings [2]. Traditional collaborative filtering system sourced recommendation data through user-item interactions or user-user interactions. Pure content based recommender system utilizes user profiles. However, both methods result in cold start which in turns lead to the RE [3]. Utilizing many features (mostly irrelevant) in recommendation process leads to the RE [4]. Sometimes the chosen recommendation algorithm is the source of RE. for instance, K-Nearest Neighbor Algorithms is one of the most widely used collaborative filtering algorithm. However, the error rate of this algorithm tend to be higher when the user rating is changed [5]. The Long Tail problem in recommender systems refers to the issue where a large number of items have very few ratings or interactions with users, making it difficult for recommendation algorithms to provide accurate predictions for these items [6]. This is a common issue in many recommendation scenarios, where the popularity of certain items dominates the entire dataset. As a result, many niche or less popular items are often overlooked by traditional recommendation algorithms, which can lead to missed opportunities for both users and businesses and hence, leads to RE. Popularity bias refers to an inherent tendency for people to be influenced or swayed by what is popular or already accepted. This can be a challenge in many fields, including research field recommendation system. Popularity bias is one of the factors affecting the integrity of recommendation result. In fact, popularity bias increases the ER [7]. The collaborative filtering technique leverages user-profile details, visited pages, and click information to ascertain a user's interests, thereby suggesting items related to their preferences. Existing collaborative filtering methods utilize both implicit and explicit features, often yielding favorable results in either classification or

*Corresponding author. Email: it@gmail.com

prediction. However, these systems commonly struggle to achieve high performance simultaneously across both measures [8].

Scholarly recommendation machines are also prone to errors due to the vast number of available published items. The abundance of information within vast scholarly databases presents a challenge in identifying potential researchers for productive collaboration [9]. A research paper recommendation system is a software tool or algorithm designed to assist researchers in discovering relevant scholarly articles based on their interests, preferences, and previous reading behavior [10]. The primary goal of such a system is to alleviate the burden of manually searching through vast repositories of research papers by providing personalized recommendations tailored to each researcher's needs. In the process of designing a research paper recommendation system, proper representation of articles is crucial for effectively capturing the essence of each paper and enabling accurate recommendations [10]. This involves identifying and extracting relevant features from the articles, which are then used as input for the recommendation algorithm. Research paper recommendation systems and expertise recommendation systems typically operate as independent entities within academic or research-oriented platforms. Research paper recommendation systems focus on suggesting relevant scholarly articles to users based on their interests, while expertise recommendation systems aim to identify and recommend experts or researchers with specific domain knowledge or expertise. The disconnect between research paper recommendation systems and expertise recommendation systems has led to inefficiencies in scholarly search processes and suboptimal recommendations [11]. Many recommendation frameworks have been proposed in the past to aid researchers in their research activities. However, very many of these works suffered from the problem called ***Recommendation Error (RE)*** where a returned list of recommended items is less or irrelevant to the researcher's interest. Overall, the goal of this research is to improve the performance of recommender systems by addressing the challenges posed by the cold start problem and leveraging the valuable information contained within research papers. By doing so, the research aims to minimize recommendation error and enhance the user experience in discovering relevant scholarly content.

2. RELATED LITERATUR

To tackle RE many work have been proposed in the past by different researchers. To deal with uncertainty in data distribution [2] proposed framework that model the uncertainty as a mixture of exponential power [3] proposed a novel designed aimed to do away RE experienced in traditional Collaborative filtering and content based approaches by combined structure of separate-training and joint-training together as basic for error minimization. Selecting only most relevant feature in recommendation processes is key factor for minimizing RE. [4] developed an application that gathered relevant features and reactions by user's while listening to the music. Utilizing only relevant features they were able to build Music recommendation system with error reduced. [5] proposed an improved KNN method aimed at reducing the error rate in the traditional KNN algorithm that occurs as a result of changes in the user's ratings. This error rate typically reflects in the final recommendations. Another cause of RE is Long Tail problem as mentioned above. Various techniques, have been proposed to address the Long Tail problem and improve the recommendation accuracy for less popular items. For instance, [6] divides the entire item-set into two parts: the head and the tail. It then proceeds to cluster only the tail items. Recommendations for the tail items are derived from the ratings within these clusters, while recommendations for the head items are based on the ratings of individual items. This strategy effectively utilizes clustering to improve recommendation accuracy, particularly for tail items, while maintaining personalized recommendations for head items. To mitigate the effects of popularity bias, [7] proposed multi-level method uses a switching approach. This involves switching from Collaborative Filtering (CF) to content-based (CB) recommendation technique when CF fails to find the target case. [8] introduced a collaborative filtering algorithm designed to address large datasets of users with symmetric purchasing patterns and repeated product purchases. To address the challenges of separating papers and expertise, and enhance the effectiveness of scholarly search and recommendation processes, there is a growing recognition of the need for greater integration between research paper and expertise recommendation systems. By bridging the gap between these systems and enabling seamless access to both relevant research papers and expertise, platforms can provide users with more comprehensive and personalized recommendations, ultimately improving the overall research experience. [11] proposed an approach that harmonized research papers and expertise recommendation in an academic domain.

3. PROPOSED METHODOLOGY

The proposed approach involves a series of implementation phases, beginning with the data acquisition phase, Data processing phase, then a documented experiment dataset.

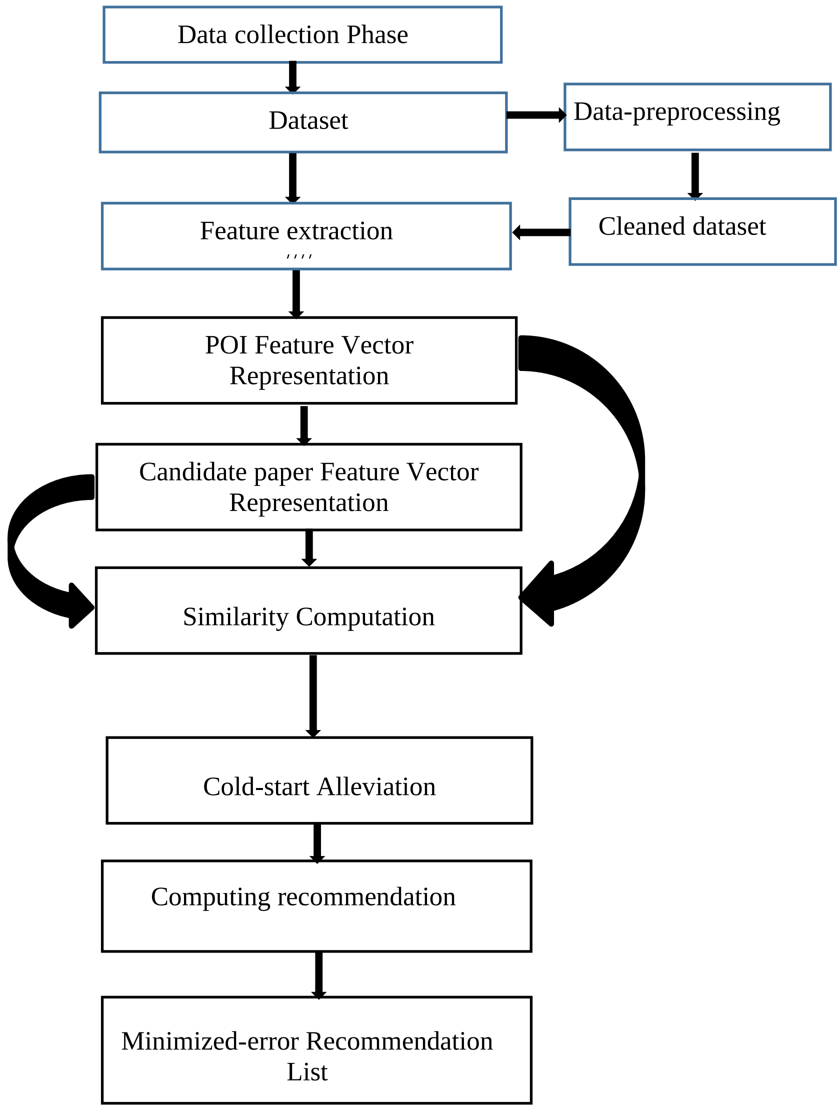

Fig .1. ACFEMSR Framework

3.1 Data Collection

When a user uploads a paper of interest (POI), relevant information such as title, abstract, keywords, and author(s) are extracted from it. However, this data needs to be processed and cleaned up in order to be useful for our purposes. This involves applying techniques such as removing stop words to make the data usable and meaningful

3.2 Content Based Vocabulary building (Feature Vecttors) and Similarity computations

In this stage feature vectors and similarity calculations are performed between the paper of interest and other potential papers based on their titles, keywords, abstracts, and authors. This involves integrating public contextual metadata, which is accomplished through a series of steps.

Step 1: Feature vector Computation for research paper F^{POI} and paper of interest (POI)

$$F^{POI} = T_{Title} + T_{Keywords} + T_{Abstract} + T_{Author(s)} \quad (1)$$

$$\text{POI Vector} = (V_1^{POI}, V_2^{POI}, V_3^{POI}, \ldots\ldots., V_n^{POI}) \quad (2)$$

Here V1 to Vn are the set of unique words found in the content of title, keywords, abstract and author(s) features of the POI.

Creating vector FCi (i = 1, 2, . . . j) for all the eligible papers. Each paper to be recommended, Ci (i = 1, 2, 3, . . . j), is represented as a feature vector FCi using the following equations:

$$F^{Ci} = \sum_{k=1}^{m} T_t + \sum_{l=1}^{n} T_K + \sum_{q=1}^{o} T_A + \sum_{r=1}^{s} T_{Author(s)} \quad (3)$$

$$= (W_{t1}^{POI}, W_{t2}^{POI}, W_{t3}^{POI}, \ldots\ldots, W_{tr}^{POI}) \quad (4)$$

3.3 Now that feature vectors of POI and candidate papers are computed, we're all set to compute the similarity between the F^{POI} and the eligible papers F^{Ci} (i=1,2,...j) .

To compute similarity between these two vectors, the cosine similarity formula is employed.

$$\boldsymbol{SIMILARITY}(\boldsymbol{F^{POI}}, \boldsymbol{F^{C}}) = \frac{\sum_{i=1}^{n} F^{POI} F^{C}}{\sqrt{\sum_{i=1}^{n} (F^{POI})^2} \sqrt{\sum_{i=1}^{n} (F^{C})^2}} = \frac{f^{POI}.f^{C}}{|f^{POI}|.|f^{C}|} \quad (5)$$

Where $\boldsymbol{f^{POI}}$ ***and*** $\boldsymbol{f^{C}}$ represent the feature vectors for the researcher's POI and qualified eligible paper

3.4 Propose IPSPR Algorithm

Algorithm 1 for Retrieving Potential Papers and Associated Public Contextual Metadata

Input: Paper (POI)

Output: Top-N recommended papers

Get a paper from a user,

(1) Get the information contained in the Title, Keywords, Abstract, and Author(s) sections of the Paper of Interest (POI)

(2) Compute feature vector F^{POI} for the POI

(3) Get all the qualified candidate papers (CP)

(4) For each CP compute feature vector F^{C}

(5) calculate the cosine similarity between POI and each of the CP

(6) list Top-N recommended papers

3.5 Experiments

3.5.1 Dataset

In this section, the Evaluation arrangement of the suggested method in order to assess its effectiveness is detailed. A publicly accessible dataset computed in [12] is utilized. The dataset includes the publications of 50 researchers in different fields. Our approach gathered publicly accessible information of their publications, such as title, abstract, keywords, and author(s). The summary of the dataset is given in Table 1 below.

TABLE I. EXPERIMENT DATASET

Table 1 Experiment Dataset	***Number of researchers***	***Average publication per researcher***	***Average citation per publication***	***Average references per publication***	***Total number of papers***	***Average citation per candidate papers***
	50	10	14.8 (Max. 169)	15.0 (Max. 58)	100,351	17.9 (Max 175)

3.6 Baseline Methods

To evaluate the performance of our approach, the experimental results were compared with the following baseline approaches:

3.6.1 Baseline 1: HPSPRPCM

A Hybrid Personalized Scientific Paper Recommendation Approach Integrating Public Contextual Metadata (HPSPRPCM) [13] presented a hybrid approach that utilizes public data (title, abstract, and keywords) in recommendation processes. This approach finds similar papers to the POI by computing the feature vectors of both POI and candidate papers. Candidate papers that has highest similarities are therefore present to the user as recommendation**.**

3.6.2 Baseline 2: RPRS

A research paper recommender system based on public contextual metadata This approach was proposed by. The authors utilize public content (title and abstract) from the papers to generate a recommendation.

Unlike the aforementioned baselines, which utilize only three (title, abstract, and keywords) and two (title and abstract) publicly available contextual metadata respectively, the proposed approach takes advantage of all four contextual metadata (title, abstract, keywords, and authors). The additional information provides more insight about papers and therefore provides a more reliable recommendation as compared with its baseline counterparts.

3.7 Evaluation Metrics

The proposed system's performance is evaluated using the metrics listed below:
MAE and RMSE: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are metrics used to evaluate a Regression Model. These metrics tell us how accurate our predictions are and, what is the amount of deviation from the actual values. Technically, RMSE is the **r**oot of the **m**ean of the **S**quare of **e**rrors and MAE is the **m**ean of **a**bsolute value of **e**rrors (Acharya, 2021). They are calculated as follows:

$$RMSE = \sqrt{\frac{\sum(y_i - y_p)^2}{n}} \tag{1}$$

$$MAE = \frac{|(y_i - y_p)|}{n} \tag{2}$$

$$y_i = Actual\ value$$
$$y_p = Predicted\ value$$
$$n = Number\ of\ observation\ /rows$$

4. RESULTS AND DISCUSSION

MAE and RMSE have been tested. The obtained experimental results from the metrics are shown in Figure 2 and Figure 3. The figures summarize the performance of each baseline approach and the proposed approach. To provide a clear understanding of the obtained results, a visual graphical representation was also displayed.

Fig 2 The performance of the proposed approach is compared with the baseline methods (HPSPRPCM) and (RPRS)) using the Mean Absolute Error (MAE) metric.

Fig3 The performance of the proposed approach is compared with the baseline methods (HPSPRPCM) and (RPRS)) using RMSE metric. **MAE and RMSE** have been tested. The obtained experimental results from the metrics are shown in Figure 2 and Figure 3. The figures summarize the performance of each baseline approach and the proposed approach. To provide a clear understanding of the obtained results, a visual graphical representation was also displayed.

TABLE II. MAE VALUES.

MAE	Mean Absolute Error	
	Approaches	***Mae Values***
1	Haruna et al (2020)	0.7
2	Sakib et al. (2021)	0.4
3	Propose Approach	0.2

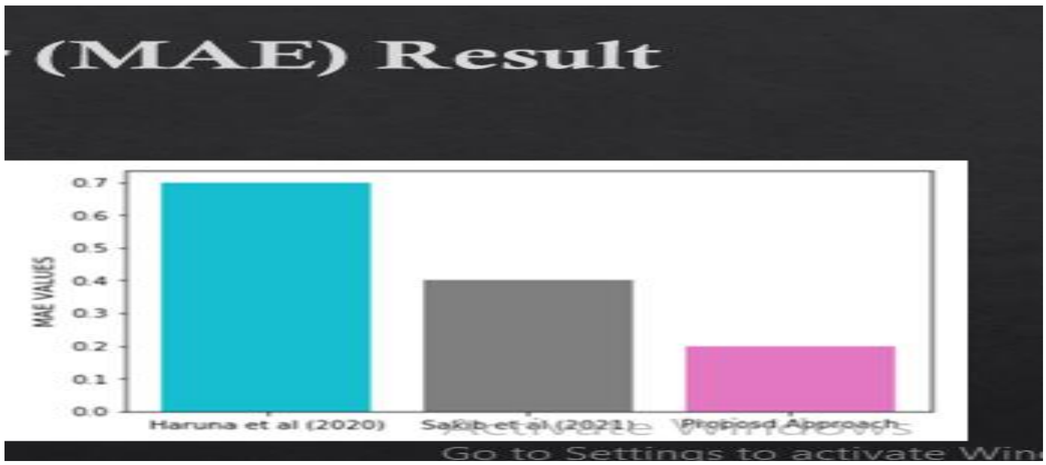

Fig. 2. HPSPRPCM and RPRS using the Mean Absolute Error (MAE) metric.

TABLE III. RMSE VALUES.

RMSE	Root Mean Square Error	
	Approaches	***RMSE Values***
1	Haruna et al (2020)	2.9
2	Sakib et al. (2021)	1.8
3	Propose Approach	0.8

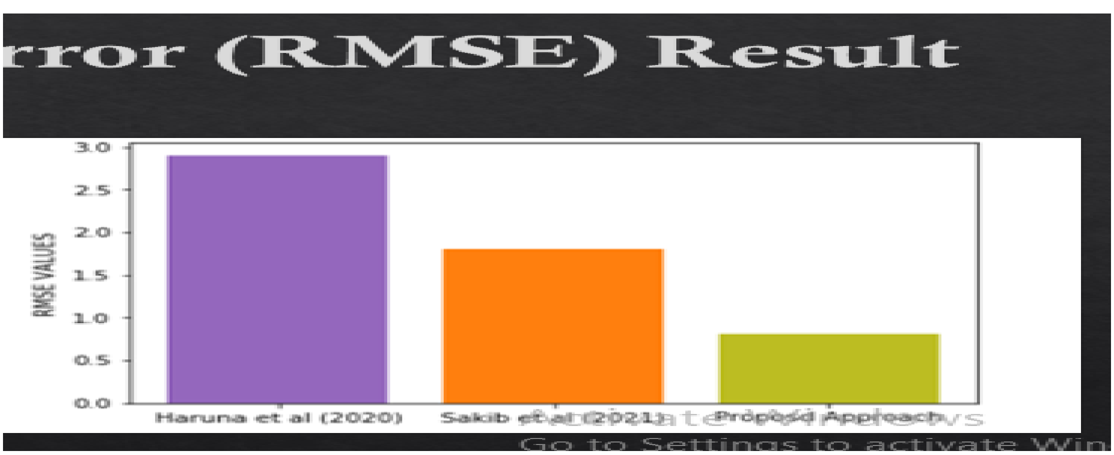

Fig. 3. HPSPRPCM and RPRS using RMSE metric.

This section presents and discusses the outcomes of the benchmark methodologies as well as the recommended methodology.

Fig. 2 and Fig. 3 above show the success rates of benchmark approaches and the suggested approach based on error measuring metrics. The graph clearly shows how the suggested method significantly reduces the error rate observed by both benchmark approaches.

5. CONCLUSION

In this paper, we introduce an author-centric framework aimed at minimizing error in scholarly recommender systems (ACFEMSR). Unlike conventional approaches, our framework incorporates the author(s) feature during the pre-recommendation process to reduce recommendation error (RE). By integrating author information, our approach enhances the system's ability to detect relevant scholarly papers, ultimately improving the accuracy of recommendations.

The objective of our research is to leverage key features in the recommendation process to lower the RE rate in scholarly recommender systems. Through experimentation using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) evaluation metrics, we demonstrate that our proposed approach effectively reduces RE compared to baseline methods. These results validate the reliability and effectiveness of our approach in improving scholarly recommendation systems.

Conflicts Of Interest

The authors declare no conflicts of interest.

Acknowledgment

The authors express their gratitude to their family and friends for their constant support and advice. We also extend our heartfelt thanks to the Head of the Department of Computer Science at UMYU for his steadfast support, encouraging words, and guidance from the beginning of this work to the completion of the manuscript.

References

[1] H.-N. Kim, A.-T. Ji, H.-J. Kim, and G.-S. Jo, "Error-Based Collaborative Filtering Algorithm for Top-N Recommendation," in *Advances in Data and Web Management*, G. Dong, X. Lin, W. Wang, Y. Yang, and J. X. Yu, Eds., Berlin, Heidelberg: Springer, 2007, pp. 594–605. doi: 10.1007/978-3-540-72524-4_61.

[2] H.-R. Zhang, Y. Qiu, K.-L. Zhu, and F. Min, "Lower bound estimation of recommendation error through user uncertainty modeling," *Pattern Recognition*, vol. 136, p. 109171, Apr. 2023, doi: 10.1016/j.patcog.2022.109171.

[3] Z. Zhu, S. Sefati, P. Saadatpanah, and J. Caverlee, "Recommendation for New Users and New Items via Randomized Training and Mixture-of-Experts Transformation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, in SIGIR '20. New York, NY, USA: Association for Computing Machinery, Jul. 2020, pp. 1121–1130. doi: 10.1145/3397271.3401178.

[4] M. Kleć, A. Wieczorkowska, K. Szklanny, and W. Strus, "Beyond the Big Five personality traits for music recommendation systems," *J AUDIO SPEECH MUSIC PROC.*, vol. 2023, no. 1, p. 4, Jan. 2023, doi: 10.1186/s13636-022-00269-0.

[5] G. Li and J. Zhang, "Music personalized recommendation system based on improved KNN algorithm," in *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Oct. 2018, pp. 777–781. doi: 10.1109/IAEAC.2018.8577483.

[6] Y.-J. Park and A. Tuzhilin, "The long tail of recommender systems and how to leverage it," in *Proceedings of the 2008 ACM conference on Recommender systems*, in RecSys '08. New York, NY, USA: Association for Computing Machinery, Oct. 2008, pp. 11–18. doi: 10.1145/1454008.1454012.

[7] G. Alshammari, J. L. Jorro-Aragoneses, N. Polatidis, S. Kapetanakis, E. Pimenidis, and M. Petridis, "A switching multi-level method for the long tail recommendation problem," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 6, pp. 7189–7198, Jan. 2019, doi: 10.3233/JIFS-179331.

[8] Z. Shahbazi, D. Hazra, S. Park, and Y. C. Byun, "Toward Improving the Prediction Accuracy of Product Recommendation System Using Extreme Gradient Boosting and Encoding Approaches," *Symmetry*, vol. 12, no. 9, Art. no. 9, Sep. 2020, doi: 10.3390/sym12091566.

[9] T. Pradhan and S. Pal, "A multi-level fusion based decision support system for academic collaborator recommendation," *Knowledge-Based Systems*, vol. 197, p. 105784, Jun. 2020, doi: 10.1016/j.knosys.2020.105784.

[10] A. Chaudhuri, N. Sinhababu, M. Sarma, and D. Samanta, "Hidden features identification for designing an efficient research article recommendation system," *Int J Digit Libr*, vol. 22, no. 2, pp. 233–249, Jun. 2021, doi: 10.1007/s00799-021-00301-2

[11] C. N. Mabude, I. O. Awoyelu, B. O. Akinyemi, and G. A. Aderounmu, "An Integrated Approach to Research Paper and Expertise Recommendation in Academic Research," *IJACSA*, vol. 13, no. 4, 2022, doi: 10.14569/IJACSA.2022.0130456

[12] K. Sugiyama and M.-Y. Kan, "Scholarly paper recommendation via user's recent research interests," in *Proceedings of the 10th annual joint conference on Digital libraries*, Gold Coast Queensland Australia: ACM, Jun. 2010, pp. 29–38. doi: 10.1145/1816123.1816129.

[13] N. Sakib *et al.*, "A hybrid personalized scientific paper recommendation approach integrating public contextual metadata," *IEEE Access*, vol. 9, pp. 83080–83091, 2021.

[14] K. Haruna, M. A. Ismail, A. B. Bichi, V. Chang, S. Wibawa, and T. Herawan, "A Citation-Based Recommender System for Scholarly Paper Recommendation," in *Computational Science and Its Applications – ICCSA 2018*, vol. 10960, O. Gervasi, B. Murgante, S. Misra, E. Stankova, C. M. Torre, A. M. A. C. Rocha, D. Taniar, B. O. Apduhan, E. Tarantino, and Y. Ryu, Eds., in Lecture Notes in Computer Science, vol. 10960. , Cham: Springer International Publishing, 2018, pp. 514–525. doi: 10.1007/978-3-319-95162-1_35.