INTERNATIONAL JOURNAL OF SOFTWARE ENGINEERING AND COMPUTER SYSTEMS
E-ISSN: 2180-0650
VOLUME 10, ISSUE 1, 2024, pp. 40 – 49
DOI: https://doi.org/10.15282/ijsecs.10.1.2024.4.0122

RESEARCH ARTICLE

# An Author-Centric Scientific Paper Recommender System to Improve Content-Based Filtering Approach

**Zaharaddeen Adamu Hamisu, Nura Mukhtar***

Faculty of Natural and Applied Science, Umaru Musa Yaradua University (UMYU), 820102 Katsina, Nigeria.

**ABSTRACT -** Scholarly publications on the web are rapidly expanding, making it difficult for scholars to identify relevant study materials. Information overload makes it harder to find important material, especially for new researchers. Scholarly recommender systems solve this issue by employing recommendation techniques to assist researchers in locating appropriate literature based on their interests. Existing systems frequently rely on user profiles and public and non-public metadata, which leads to the persistent problem in scholarly recommendations called cold start. To deal with the challenges of cold start in scholarly-based recommender systems, this research suggests an improved Content-Based Filtering (CBF) approach that takes advantage of publicly available metadata, specifically the author(s) feature. The approach incorporated author(s) features into a scholarly recommender system to serve as a basis and key component for alleviating "A New Paper Cold Start Problem." The proposed approach implements the feature vectors of the metadata using the Count Vectorizer and similarity computation was performed using the Cosine Similarity formula." An experiment using a publicly available dataset shows that the suggested method surpasses the approaches previously proposed by other researchers regarding recommendation accuracy and relevancy, making it a dependable and efficient instrument for scholarly recommendation. The result also shows the effectiveness of the author(s) feature in tackling new papers in scholarly recommendation systems.

**ARTICLE HISTORY**
Received : 21 March 2024
Revised : 23 August 2024
Accepted : 13 September 2024
Published : 20 September 2024

**KEYWORDS**
*Scientific paper*
*Recommendation*
*Public contextual metadata*
*Content-based filtering*
*Cold start recommendation*
*Author feature*

## 1.0 INTRODUCTION

The diversity of information on the net and its daily increases have posed challenges to anyone looking for a particular item of information. According to Mohanty *et al.* [1] Roughly 25 million academic papers are available for access via the Internet. This phenomenon presents the difficulty of dealing with an excessive amount of information- where researchers encounter a multitude of search results, with the majority often being irrelevant to their queries [2]. The need to manually filter hundreds or even thousands of pieces of information looking for an item of interest is not just time-consuming but also leads to choosing unrelated or irrelevant data for our interests. To solve this issue, the current research suggests an enhanced scholarly recommendation framework that utilizes publicly available metadata from the research paper to recommend an item to researchers. Additionally, the new paper cold start is tackled using the author(s) feature.

Many systems are currently available to aid researchers in their work, for example, consider Google Scholar, DBPL, and Xueshu Baidu. However, working with these systems is not as simple as it may seem because the many results matching the scholar's request are not important [3]. Search engines, like Google Scholar and Microsoft Academic Search, make recommendations utilizing keywords provided by the researchers. The keyword-based search can produce very confusing outcomes as keywords are not enough to represent the actual papers needed by the researcher Dai *et al.* [4].

The challenges mentioned above are tackled using a smart system (a recommendation system) that helps users in their search. A recommendation system is a technology adopted in many fields that involves the selection of items from a repository of similar items, such as movies, shopping, news, scholarly research publications, and so on. Working out recommendation machines for a scholarly recommendation has attracted many researchers Sakib *et al.*[5]. Many works are done using different approaches by many researchers. In a nutshell, the scholarly recommender system is designed in three main ways: Content-Based Filtering (CBF), collaborative Filtering (CF), and a hybrid approach.

The CBF analyses the information contained in the research paper, including keywords, abstract, title, authors, introduction, etc. It then mines the information to construct the researcher's profile. For instance, [6] built a content-based recommendation system that manages common author relations between articles. Here, two articles are considered to be related or similar if they have the same author(s). [7] introduced a research paper recommendation approach that employs

***CORRESPONDING AUTHOR |N. Mukhtar** | ✉ nura.mukhtar@umyu.edu.ng

the researcher's prior works to develop their profile. In a study [8] a web crawler extracts scientific articles from the internet. Then similarity between the two research papers based on their textual similarities is determined. Metadata has been proven to be very important in scholarly recommendation systems. Using relevant metadata both researcher's profile and articles can be explored as In [9]. The main problem associated with pure CBF scholarly approaches is that they require the researcher's profile to make recommendations.

To address the shortcomings of content-based filtering (CBF), the research community has shown significant interest in Collaborative Filtering (CF) as an alternative recommendation approach. As an example, Haruna *et al.* [10] introduced a recommender system that relies on citations, utilizing hidden connections among research papers to suggest scholarly articles. Boussaadi *et al*. [11] exploit the relevance of the research community through the thematic structure of the research community to generate personalized recommendations for a given user. collaborative filtering (CF) method is very effective especially when the paper content is not needed. Amami et al. [12] suggested a graph-based technique in which the researcher's profile is built based on the topic they have visited or rated. Manju *et al.* [13] introduced a random walk strategy on a diverse paper graph that uses a Mendeley social network to make recommendations and, at the same time, alleviate cold start. In the collaborative filtering (CF) approach to scholarly recommendation, authors are treated as items to be rated, and activities such as co-authoring papers are regarded as rating actions Zhang *et al.* [14]. The whole idea of CF lies in the collaborators with whom a researcher associates. Young or novice researchers find it very difficult to collaborate with other researchers or make a good relevant collaboration. Like CBF, CF also has several drawbacks. One of those drawbacks is the difficulty in finding a suitable rating for the given paper. Hence, most recommendation systems that utilize a single approach, like CBF or CF, have individual drawbacks.

Hybrid recommender systems were intended to overcome the shortcomings experienced in the CBF and CF, hence they have been utilized in the scholarly recommendation system these days. Hybrid makes use of CBF and CF features simultaneously to take advantage of each other. [15] proposed a citation-based approach that leverages the power of the web of data to build a recommendation engine. The methodology considers a substantial volume of data gathered from diverse platforms, including those in the realm of social media. [5] proposed a hybrid paper recommendation method that makes use of multi-level citation networks and public metadata in paper recommendations. [16] suggested a hybrid Scholarly recommendation model that incorporates contextual metadata, including titles, keywords, and abstracts, with conventional Content-Based Filtering (CBF) to discern contextual similarities among papers. The model also utilizes contextual citation relations with traditional CF to find citation relations.

Despite the robustness of the proposed system, there is a need to incorporate additional information, such as authors, into the content-based part of the system to increase its accuracy. This gap means that the proposed approach has not yet solved the new paper's cold-start problem. For instance, an author may have published a new paper, but it's yet to be cited by anyone; therefore, this newly published paper cannot be listed in the candidate papers of the paper of interest using this method. Hence, this research work aims to overcome the cold start challenge of the newly published paper by incorporating the authors of the paper among the futures in the system. By incorporating the author(s) feature into the recommendation process, the system can now detect and discover related papers authored by the same researchers. This integration significantly enhances the system's ability to recommend newly published papers by leveraging the author's previous work and contributions to the field. The system ranks the candidate papers based on their similarity scores and generates a list of the Top recommended papers. This approach ensures that newly published papers, which might suffer from the cold start problem due to insufficient metadata, can still be effectively recommended by leveraging the additional author(s) feature. The integration of this feature enriches the recommendation process, leading to more accurate and relevant suggestions for users.

The main achievements of this work are listed below:

1. An improved personalized scholarly-based recommendation (IPSPR) is proposed that utilizes only freely accessible paper information.
2. A scholarly recommendation application that overcomes the cold start problem of the new paper by incorporating the author feature of the research paper in the recommendation process.

The remaining part of the paper is arranged in the following structure: Section two reviews related works on previous research paper recommendation approaches and cold-start alleviating approaches. Section three details our approach, while Section four covers the Setting up of the experiment and assessment measures. The findings are examined and debated in Section Five. Lastly, section Six provides the conclusion of the paper.

## 2.0 RELATED WORKS

Recommender Systems (RSS) is a burgeoning research field that has gained significant attention from researchers due to its wide-ranging importance, leading to rapid and substantial growth. The daily increase in researchers' interest in the field is proportional to the overwhelming increase in internet technology and E-businesses. This Part of the paper highlights some of the previous methods used by the researchers in developing the Recommender system.

Content-based filtering (CBF) is among the dominant approaches used for the development of recommendation systems. It utilizes the features or descriptions of an item to create a user profile according to the item's feature. To recommend new items to the user using this method, previous users' activities are analyzed to find a pattern through which similarities are established Koren *et al.* [17]. [18] built a recommendation system that manages common author relations between articles. Here, two articles are considered to be related or similar if they have the same author(s). Finally, the Random Walk algorithm is implemented to compute similarities and obtain the top articles to recommend to a researcher.

To integrate the recommendation system into the digital library, the term Frequency-Inverse Document Frequency (TF-IDF) and cosine similarity have been employed to figure out the connection an article of research corresponds to a user's search or topic of concern. An example of the work that combined TF-IDF and cosine similarity was a framework proposed by Lofty *et al.* [21] that was based on adaptive hypermedia. [22] also takes advantage of the co-authorship network to construct the research profile using TF-IDF. Insufficient information always minimizes the accuracy of any recommender system. For instance, in a situation where a recommender is designed to use both or sometimes only non-public datasets, the system may run into problems when this data is not available. Hence, using publicly available paper information or metadata makes the best recommendation system as proposed by Zhu *et al.* [23].

Chi *et al.* [19] proposed a keyword-based concept recommendation model to remind users of related input keywords and expand their search terms. The proposed system generates concept terms related to the search term the user presented. Hong *et al.* [20] proposed a user-centric profile scholarly recommender system that uses an algorithm to extract keywords from the paper. That is, the work relied heavily on the keyword metadata, which is not enough attributes for the recommendation process.

Unlike many recommendation approaches that treated the whole document as a single topic, [24] proposed a content-based Recommendation approach that considered an article as a constituent of various topics through which two more articles' similarities can be deduced. Traditional Content-based recommender system normally utilizes the user's profile taking into account His/her previous publications to make recommendations for new articles.

The cold start describes the difficulty in making a recommendation when the item is new. The difficulty remains a persistent problem for the CF recommender system. [25] divided answers that addressed the cold start dilemma into ten groups according to the technique and methods adopted. In the case of insufficient details concerning the individual or items, the recommendation engine is adjusted to source the required information from an external source, such as social networking sites. The sourced information is now used to bootstrap the system to do away with the cold start issues earlier encountered by the system. One such work that implements such idea of utilizing available social network data was proposed by Shaphira *et al.* [26]. [27] proposed a cold start alleviating approach that populates cold user profiles by finding the relationship among the users through mining the pattern of their research works. Similarly, Shaw *et al.* [28] proposed a framework that expands the profiles of the researchers by finding and computing the pattern of association in the dataset. Lika *et al.* [29] the method of classification is incorporated into the pure CF, and demographic information is utilized to find users with similar interests. A covering reduction algorithm was proposed by Zhang *et al.* [30] which removes redundant users from around the new user to avoid the cold start.

In [31] a method to tackle a cold start problem in the scholarly recommender system was proposed. The solution utilized the individual's networking relationships and subject interactions among publications in the Mendeley scholarly online communities. In [32, 5] citation networks were utilized to discover many papers hidden in citation relations. New academicians and new papers are mostly undetected by the recommender system (papers and researchers cold start).

To overcome the issue of cold start, Gupta *et al.* [33] proposed a framework that provides a rating for each work done by the researcher and established a concealed relationship that connects the paper's title, references, and citations. To do away with the problem of item cold start suffered by Micro Open Education Resources (OREs) a Heuristic approach is proposed by Sun *et al.* [34] to incorporate recently released OREs into the existing training path to increase the chances of paper discovery. Citations analysis and networks were used by researchers to fix the cold start difficulty. For instance, the similarities of articles could be discovered by identifying the rhetorical zone of each article, Abbas *et al.* [35]. However, this work utilizes only the Abstract and introduction content of the paper, and this is insufficient metadata for recommendation. Subsequently, a cold start problem is dynamic and could be addressed using different methods. Hence, this work aims to tackle the new paper cold start using Metadata of the paper specifically the author(s) metadata feature.

## 3.0 METHODS AND MATERIAL

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

### 3.1 Data Acquisition

The user submits a paper of interest (POI) to the system. Every piece of the available public contextual information for the selected paper of interest is extracted, which includes the title, keywords, abstract, and author(s). The data obtained is unusable in its actual format, therefore it has to undergo a wrangling process to make it clean and usable for our purpose. Cleaning processes such as removing the stop words are done to the data.

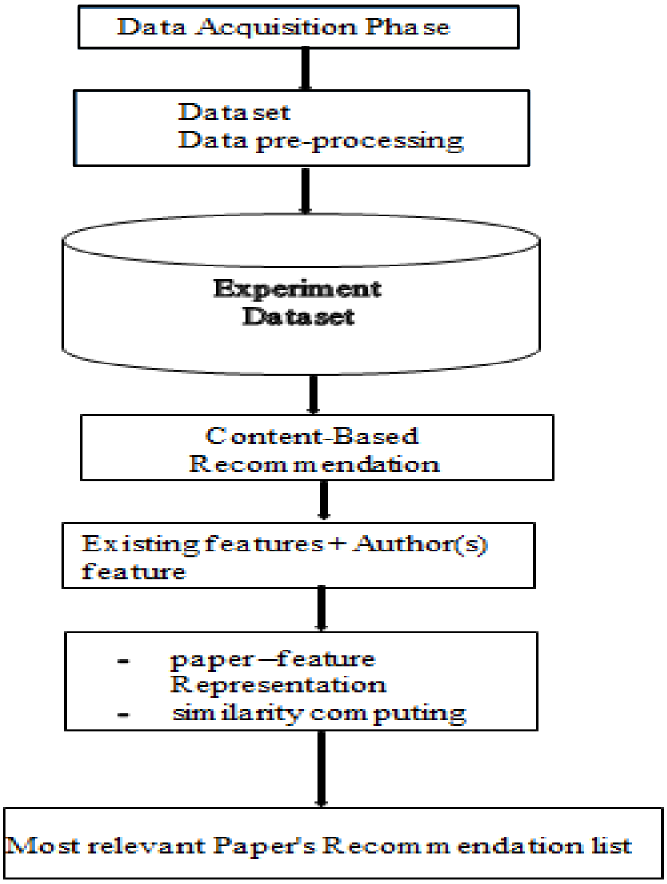


Figure 1: IPSPR Flowchart

### 3.2 The Content-Based Model Utilizing Public Contextual Metadata

This phase performs similarity computation between POI and candidates' papers based on Title, keywords, abstract, and author(s). The integration of public contextual metadata involves the following steps:

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

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

$$= (W_{t1}^{POI}, W_{t2}^{POI}, W_{t3}^{POI}, \ldots\ldots, W_{tr}^{POI}) \quad (2)$$

In this context, 't' is the difference between the terms found in the Title, Keywords, Abstract, and Author(s) of an article of interest. Additionally, 'W' denotes each term. We also need to compute the feature vector for the qualified candidate papers. For each selected paper, construct the feature vector $F^{C}$ (i = 1, 2,...j). Every potential article to recommend Ci (i = 1, 2, 3,...j) is identified as a feature vector $F^{Ci}$.

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

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

Now that the feature vectors of POI and candidate papers are computed, we're all set to compute the similarity between the feature vector of the paper of interest and the potential paper (i=1,2,...j).

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

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

Where $f^{POI}$ *and* $f^{C}$ denote the feature vectors for the researcher's POI and potential candidate papers.

Propose IPSPR Algorithm:

Algorithm 1: IPSPR Algorithm for Retrieving Potential Articles and Associated Available Information

Input: Paper of Interest (POI)

Output: Top-N recommended papers

1. Retrieve Information from Paper of Interest (POI):
   - Extract the title, keywords, abstract, and author(s) information from the Paper of Interest (POI).
   - Data cleaning is also performed at this stage using pandas to obtained a practicable dataset.
2. Compute Feature Vector $f^{POI}$ for POI:
   - Generate a feature vector ($f^{POI}$) representing the semantic content of the POI.
3. Retrieve Qualified Candidate Papers (CP):
   - Query the database or dataset to retrieve a set of selected papers (CP) that are potentially related to the POI.
4. Compute Feature Vectors for Candidate Papers:
   - Generate feature vectors ($f^{C}$) for each candidate paper (CP) by processing the textual content of their titles, keywords, abstracts, and author(s) information.
5. Calculate Cosine Similarity:
   - Compute the cosine similarity between the feature vector of the POI ($f^{POI}$) and each candidate paper's feature vector ($f^{C}$) to measure their semantic similarity.
6. Select Top-N Recommended Papers:
   - Rank the candidate papers based on their cosine similarity scores and select the Top-N papers as the recommended papers for the user.

**3.3 Steps of the algorithm and implementation of the study**

1. Retrieve Information from Paper of Interest (POI):
   - Extract the title, keywords, abstract, and author(s) information from the Paper of Interest (POI).
   - Data cleaning is also performed at this stage using pandas to obtained a practicable dataset.
   - Once the information is extracted, the next step is to clean the data using pandas, a popular data manipulation library in Python. Data cleaning ensures that the dataset is free from errors, inconsistencies, and irrelevant data, making it more useful for analysis or further processing.
2. Compute Feature Vector $f^{POI}$ for POI:
   - Generate a feature vector ($f^{POI}$) representing the semantic content of the POI.
   - The feature vector is generated using The CountVectorizer
   - The CountVectorizer converts text into a numerical format that machine learning algorithms can use. Each document is represented as a vector of word counts, allowing algorithms to analyze and model the text data effectively.
3. Retrieve Qualified Candidate Papers (CP):
   - Query the database or dataset to retrieve a set of selected papers (CP) that are potentially related to the POI.
4. Compute Feature Vectors for Candidate Papers:
   - Generate feature vectors ($\boldsymbol{f^{C}}$) for each candidate paper (CP) by processing the textual content of their titles, keywords, abstracts, and author(s) information.
   - The CountVectorizer converts text into a numerical format that machine learning algorithms can use. Each document is represented as a vector of word counts, allowing algorithms to analyze and model the text data effectively.
5. Calculate Cosine Similarity:
   - Compute the cosine similarity between the feature vector of the POI ($f^{POI}$) and each candidate paper's feature vector ($\boldsymbol{f^{C}}$) to measure their semantic similarity.

- Once the text is vectorized, you can compute the cosine similarity between the two vectors (representing the POI and the Candidate Paper). Scikit-learn's cosine_similarity function computes this similarity score.

6. Steps to Select Top-N Recommended Papers: Compute Cosine Similarity:

- First, as explained earlier, compute the cosine similarity scores between the POI and each candidate paper. These scores reflect how similar each candidate paper is to the POI, with higher scores indicating greater similarity. Rank the Candidate Papers:
- Once the cosine similarity scores are computed, the next step is to rank the candidate papers. This ranking is done in descending order, meaning that the paper with the highest similarity score is ranked first.

### 3.4 Overview of The Proposed Model.

The proposed approach enhances the scholarly recommendation system by effectively improvement of the accuracy of system. This is achieved through the integration of the author(s) feature along with the title, keywords, and abstract. The model leverages a Content-Based Recommendation (CBR) approach to compute the similarity between the Paper of Interest (POI) and candidate papers using their contextual metadata. The model operates by first receiving a query from a user, which is the title of the Paper of Interest (POI). The system then extracts required metadata from the POI, specifically the title, keywords, abstract, and author(s). Using this metadata, a feature vector for the paper of interest ($F^{POI}$) is computed. Next, the system retrieves a set of qualified candidate papers (CP). For each candidate paper, a feature vector ($F^{FC}$) is computed using the same types of metadata extracted from the POI. The similarity between the POI and each candidate paper is then calculated using the cosine similarity measure, which determines how closely related the papers are based on their feature vectors.

Finally, the system ranks the candidate papers based on their similarity scores and generates a list of the most relevant recommended papers. This approach ensures that newly published papers, which might suffer from the cold start problem due to insufficient metadata, can still be effectively recommended by leveraging the additional author(s) feature. The integration of this feature enriches the recommendation process, leading to more accurate and relevant suggestions for users.

### 3.5 EXPERIMENTS

#### 3.5.1 Dataset

In this section, the Evaluation arrangement of the suggested method to assess its effectiveness is detailed. A publicly accessible dataset computed in [24] is utilized. The dataset includes the publications of 50 researchers in different fields. Our approach gathered publicly accessible information about every one of their articles, including the title, abstract, keywords, and author(s). Table 1 presents an overview of the dataset.

Table 1. Experiment Dataset

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

#### 3.5.2 Benchmark approaches

To evaluate the performance of Our technique, outcomes from experiments were compared with the following baseline approaches:

##### 3.5.2.1 Hybrid Personalized Scientific Paper Recommendation Approach Integrating Public Contextual Metadata (HPSPRPCM)

A Hybrid Personalized Scientific Paper Recommendation Approach Integrating Public Contextual Metadata (HPSPRPCM) [5] presented a hybrid approach that utilizes public contextual metadata (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 with the highest similarities are therefore suggested to the user.

#### 3.5.2.2 Research Paper Recommender System (RPRS)

A research paper recommender system based on public contextual metadata This approach was proposed in [36]. The approach utilized available content (title and abstract) from articles 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 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.

### 3.6 Evaluation Metrics

To determine the effectiveness of the suggested framework, the following metrics are used, which are widely used in evaluating recommender systems:

#### 3.6.1 Precision

Precision is the fraction of correctly recommended papers out of all recommended papers. Written as a formula:

$$\text{Precision (P)} = \frac{\text{X}}{\text{Y}} \tag{6}$$

Where X denotes Number of papers correctly Recommended and Y denotes Total Recommended Papers.

#### 3.6.2. Recall

Recall is the fraction of all relevant papers returned by the system. Written as a formula:

$$Recall\ (R) = \frac{\text{Ri}}{\text{Rt}} \tag{7}$$

where Ri denotes Number of papers correctly recommended and Rt denotes Number of papers correctly recommended

#### 3.6.3 F1

The F1 harmonizes precisions and recall to obtained better understanding of the performance of the system.

$$F1 = \frac{2*\text{P}*\text{R}}{\text{P}+\text{R}} \tag{8}$$

Where P and R, denotes Precision and Recall respectively.

#### 3.6.4 Mean Average Precision (MAP)

Mean Average Precision is used to measure the overall effectiveness of the search algorithm.

$$MAP = \frac{1}{\text{n}}\sum_{k=1}^{n} AP_{\text{k}} \tag{9}$$

Where n is the number of average precisions in a set and $AP_k$ is the average precision (AP) for a given query.

## 4.0 RESULTS AND DISCUSSION

Outcomes of the benchmark approaches and suggested approaches are presented and discussed in this section. Figure 2 shows the effectiveness of the benchmark methods and the suggested methods as per the precision metric. The graph provides a clear demonstration of how the proposed approach markedly excelled over the benchmark methods in precision values at all K-values. For instance, when K = 15, the proposed approach has a precision value of 0.9% whereas the benchmark approach scores 0.67% and 0. 15% respectively. Considering the outcomes, we can conclude that the suggested work returns the highest percentage of relevant papers. As in precision, in recall also, the proposed approaches overtake the benchmark approaches at all values of K. However, when K = 5 both Sakib et al. (2021) and the proposed approach score 1% recall values as indicated in Figure 3. To summarize the concept, the proposed approach can identify more relevant papers to the POI compared to its benchmark counterparts, thus having the highest recall value in all scenarios (K@10, K@15, K@20, K@25, and K@30) except for K@5, where both [5] and the proposed approach have the same recall value of 1. The proposed approach demonstrated a significant performance advantage over benchmark methods based on the F1 metric. Figure 4 below depicts the performance of the three approaches based on F1. Figure 5 shows the MAP results of each benchmark approach and suggested work. According to the MAP results, the proposed approach proves its effectiveness in returning the highest percentage of the relevant papers (96%) compared with the 0.74% and 0.5% returned by [5], [33] respectively, the proposed approach has an additional advantage.

Precision

| APPROCHES | K at N | | | | | |
|---|---|---|---|---|---|---|
| K-VALUES | 5 | 10 | 15 | 20 | 25 | 30 |
| HARUNA et al (2020) | 0.6 | 0.3 | 0.2 | 0.15 | 0.12 | 0.1 |
| SAKIB et al (2021) | 1 | 0.6 | 0.53 | 0.67 | 0.72 | 0.76 |
| PROPOSED APPROACH | 1 | 0.9 | 0.866 | 0.9 | 0.92 | 0.93 |

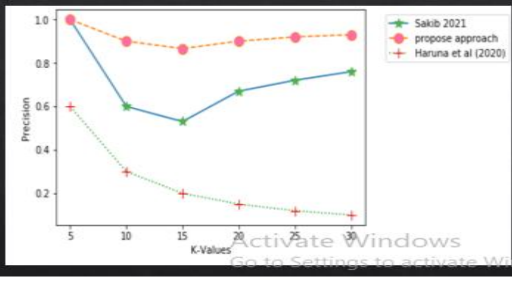


Figure 2. Compare the performance of the suggested method with two other approaches, HPSPRPCM and RPRS, in terms of Precision.

Recall

| APPROCHES | K at N | | | | | |
|---|---|---|---|---|---|---|
| K- Values | 5 | 10 | 15 | 20 | 25 | 30 |
| HARUNA et al (2020) | 0.60 | 0.25 | 0.17 | 0.13 | 0.11 | 0.09 |
| SAKIB et al (2021) | 1 | 0.5 | 0.47 | 0.56 | 0.69 | 0.72 |
| PROPOSED APPROACH | 1 | 0.75 | 0.76 | 0.78 | 0.88 | 0.88 |

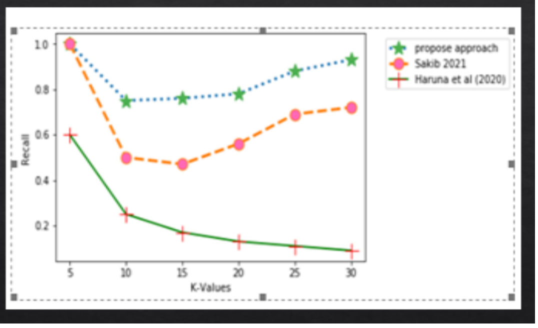


Figure 3. The performance of the suggested approach is compared with two baseline methods, HPSPRPCM and RPRS, about Recall.

F1

| APPROCHES | K at N | | | | | |
|---|---|---|---|---|---|---|
| K-VALUES | 5 | 10 | 15 | 20 | 25 | 30 |
| HARUNA et al (2020) | 0.60 | 0.27 | 0.19 | 0.14 | 0.001 | 0.017 |
| SAKIB et al (2021) | 1 | 0.49 | 0.75 | 0.61 | 0.7 | 0.74 |
| PROPOSED APPROACH | 1 | 0.82 | 0.82 | 0.84 | 0.9 | 0.91 |

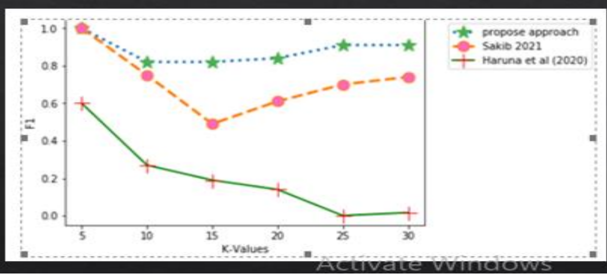


Figure 4. Comparative performance of the proposed method against the baseline approaches (HPSPRPCM and RPRS) in terms of F1 score.

Mean Average Precision (MAP) Result

| APPROCHES | MAP (FOR ALL K-VALUES) |
|---|---|
| HARUNA et al (2020) | 0.5 |
| SAKIB et al (2021) | 0.74 |
| PROPOSED APPROACH | 0.96 |

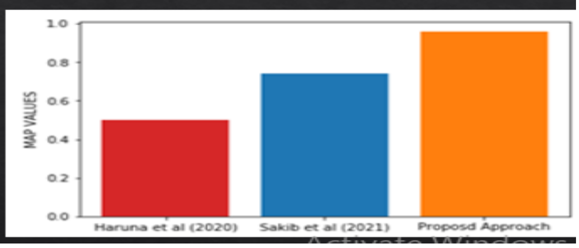


Figure 5. The proposed method compares performance with two baseline approaches, HPSPRPCM and RPRS, about Mean Average Precision (MAP).

## 5.0 CONCLUSIONS

This study, we present a paper on cold start alleviation techniques to enhance HPSPRPCM, ensuring the detection of all relevant papers published by the author(s). Unlike benchmark approaches, this work integrates the author(s) feature in the pre-recommendation phase to address the cold start problem. The integration enhanced the scholarly recommendation system's relevant papers detection capacity. In concluding remarks, in this work, an Improved HPSPRPCM Technique was proposed called IPSPR. The goal is to integrate additional features (author(s)) into the existing ones (title, abstract, and keywords) to enhance HPSPRPCM. The IPSPR is designed to be more effective and efficient only for the Mitigation of the cold start issue in the scholarly recommendation machines. For this process, the author(s) feature is considered the best feature for detecting the hidden papers published by the author. According to the experiments conducted using about four performance evaluation metrics, the outcome indicates that the suggested IPSPR successfully overtakes the benchmark approaches across the whole set of metrics. For instance, using MAP, the proposed approach scored 96%, while benchmark approaches scored 68% and 50%, respectively. The summary of the MAP result proved that the proposed approaches produce corrected recommendations in comparison with the benchmark methods. Thus, the suggested approach is more reliable.

## ACKNOWLEDGEMENTS

The authors would like to thank their family and friends for their persistent support and advice. We also extend special thanks to the Head of the Department of Computer Science at UMYU for his unwavering support, words of encouragement, and guidance from the inception of this work to the final draft of this manuscript.
This study was not supported by any grants from funding bodies in the public, private, or not-for-profit sectors.

## AUTHORS CONTRIBUTION

N. Mukhtar (Conceptualization; Formal analysis; Visualisation; Supervision; Writing - review & editing)

A.H,Zaharadeen (Methodology; Data curation; Writing - original draft; Resources)

### CONFLICT OF INTEREST

We the authors declare no conflicts of interest.

### REFERENCES

[1] S. N. Mohanty, J. M. Chatterjee, S. Jain, A. A. Elngar, and P. Gupta, *Recommender system with machine learning and artificial intelligence: practical tools and applications in medical, agricultural and other industries*. John Wiley & Sons, 2020.
[2] R. Habib and M. T. Afzal, "Sections-based bibliographic coupling for research paper recommendation," *Scientometrics*, vol. 119, no. 2, pp. 643–656, May 2019, doi: 10.1007/s11192-019-03053-8.
[3] J. Chen and Z. Ban, "Academic Paper Recommendation Based on Clustering and Pattern Matching," in *Artificial Intelligence*, vol. 1001, K. Knight, C. Zhang, G. Holmes, and M.-L. Zhang, Eds., in Communications in Computer and Information Science, vol. 1001. , Singapore: Springer Singapore, 2019, pp. 171–182. doi: 10.1007/978-981-32-9298-7_14.
[4] T. Dai, L. Zhu, Y. Wang, H. Zhang, X. Cai, and Y. Zheng, "Joint model feature regression and topic learning for global citation recommendation," *IEEE Access*, vol. 7, pp. 1706–1720, 2018.
[5] N. Sakib *et al.*, "A hybrid personalized scientific paper recommendation approach integrating public contextual metadata," *IEEE Access*, vol. 9, pp. 83080–83091, 2021.
[6] F. Xia, H. Liu, I. Lee, and L. Cao, "Scientific article recommendation: Exploiting common author relations and historical preferences," *IEEE Trans. Big Data*, vol. 2, no. 2, pp. 101–112, 2016.
[7] 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*, 2010, pp. 29–38.
[8] J. Lee, K. Lee, and J. G. Kim, "Personalized Academic Research Paper Recommendation System," arXiv.org. Accessed: Apr. 25, 2024. [Online]. Available: https://arxiv.org/abs/1304.5457v1
[9] M. Kaya, Ş. Birinci, J. Kawash, and R. Alhajj, Eds., *Putting Social Media and Networking Data in Practice for Education, Planning, Prediction, and Recommendation*. in Lecture Notes in Social Networks. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-33698-1.
[10] K. Haruna and M. A. Ismail, "Research paper recommender system evaluation using collaborative filtering," in *AIP conference proceedings*, AIP Publishing, 2018.
[11] S. Boussaadi, H. Aliane, O. Abdeldjalil, D. Houari, and M. Djoumagh, "Recommender systems based on detection community in academic social network," in *2020 International Multi-Conference on: "Organization of Knowledge and Advanced Technologies"(OCTA)*, IEEE, 2020, pp. 1–7.
[12] M. Amami, G. Pasi, F. Stella, and R. Faiz, "An LDA-based approach to scientific paper recommendation," in *Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural*

*Language to Information Systems, NLDB 2016, Salford, UK, June 22-24, 2016, Proceedings 21*, Springer, 2016, pp. 200–210.
[13] G. Manju, P. Abhinaya, M. R. Hemalatha, and G. G. Manju, “Cold start problem alleviation in a research paper recommendation system using the random walk approach on a heterogeneous user-paper graph,” *Int. J. Intell. Inf. Technol. IJIIT*, vol. 16, no. 2, pp. 24–48, 2020.
[14] Z. Zhang *et al.*, “Scholarly recommendation systems: a literature survey,” *Knowl. Inf. Syst.*, pp. 1–46, 2023.
[15] F. Zarrinkalam and M. Kahani, “A multi-criteria hybrid citation recommendation system based on linked data,” in *2012 2nd International eConference on Computer and Knowledge Engineering (ICCKE)*, IEEE, 2012, pp. 283–288.
[16] S. Hui, M. Wei, Z. XiaoLiang, J. JunYan, L. YanBing, and C. ShuJuan, “A hybrid paper recommendation method by using heterogeneous graph and metadata,” in *2020 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2020, pp. 1–8.
[17] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
[18] Y. Chi, J. Zhu, and H. Xu, “Concepts recommendation for searching scientific papers,” *Clust. Comput.*, vol. 22, pp. 8669–8675, 2019.
[19] K. Hong, H. Jeon, and C. Jeon, “UserProfile-based personalized research paper recommendation system,” in *2012 8th International Conference on Computing and Networking Technology (INC, ICCIS, and ICMIC)*, IEEE, 2012, pp. 134–138.
[20] M. M. Lofty, A. A. Salama, H. A. El-Ghareeb, and M. A. El-dusky, "Subject recommendation using Ontology for computer science ACM curricula," *Int. J. Inf. Sci. Intell. Syst.*, vol. 3, no. 1, pp. 199–205, 2014.
[21] K. Sugiyama and M.-Y. Kan, “A comprehensive evaluation of scholarly paper recommendation using potential citation papers,” *Int. J. Digit. Libr.*, vol. 16, no. 2, pp. 91–109, 2015.
[22] J. Zhu, B. G. Patra, and A. Yaseen, “Recommender system of scholarly papers using public datasets,” *AMIA Summits Transl. Sci. Proc.*, vol. 2021, pp. 672–679, May 2021.
[23] M. Ostendorff, "Contextual Document Similarity for Content-based Literature Recommender Systems." arXiv, Aug. 01, 2020. doi 10.48550/arXiv.2008.00202.
[24] V. R. Revathy and S. P. Anitha, “Cold Start Problem in Social Recommender Systems: State-of-the-Art Review,” in *Advances in Computer Communication and Computational Sciences*, vol. 759, S. K. Bhatia, S. Tiwari, K. K. Mishra, and M. C. Trivedi, Eds., in Advances in Intelligent Systems and Computing, vol. 759. , Singapore: Springer Singapore, 2019, pp. 105–115. doi: 10.1007/978-981-13-0341-8_10.
[25] B. Shapira, L. Rokach, and S. Freilikhman, "Facebook single and cross-domain data for recommendation systems," *User Model. User-Adapt. Interact.*, vol. 23, pp. 211–247, 2013.
[26] H. Sobhanam and A. K. Mariappan, “Addressing cold start problem in recommender systems using association rules and clustering technique,” in *2013 International Conference on Computer Communication and Informatics*, IEEE, 2013, pp. 1–5.
[27] G. Shaw, Y. Xu, and S. Geva, “Using Association Rules to Solve the Cold-Start Problem in Recommender Systems,” in *Advances in Knowledge Discovery and Data Mining*, M. J. Zaki, J. X. Yu, B. Ravindran, and V. Pudi, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2010, pp. 340–347. doi: 10.1007/978-3-642-13657-3_37.
[28] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, “Facing the cold start problem in recommender systems,” *Expert Syst. Appl.*, vol. 41, no. 4, pp. 2065–2073, 2014.
[29] Z. Zhang, Y. Zhang, and Y. Ren, “Employing neighborhood reduction for alleviating sparsity and cold start problems in user-based collaborative filtering,” *Inf. Retr. J.*, vol. 23, no. 4, pp. 449–472, Aug. 2020, doi: 10.1007/s10791-020-09378-w.
[30] D. Gupta, A. Nainwal, and B. Pant, “Cold Start Problem Resolution Using Bayes Theorem,” in *Research in Intelligent and Computing in Engineering*, R. Kumar, N. H. Quang, V. Kumar Solanki, M. Cardona, and P. K. Pattnaik, Eds., in Advances in Intelligent Systems and Computing. Singapore: Springer, 2021, pp. 893–899. doi: 10.1007/978-981-15-7527-3_85.
[31] G. Sun, T. Cui, D. Xu, J. Shen, and S. Chen, “A Heuristic Approach for New-Item Cold Start Problem in Recommendation of Micro Open Education Resources,” in *Intelligent Tutoring Systems*, R. Nkambou, R. Azevedo, and J. Vassileva, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 212–222. doi: 10.1007/978-3-319-91464-0_21.
[32] M. A. Abbas, S. Ajayi, M. Bilal, A. Oyegoke, M. Pasha, and H. T. Ali, "A deep learning approach for context-aware citation recommendation using rhetorical zone classification and similarity to overcome the cold-start problem," *J. Ambient Intell. Humaniz. Comput.*, May 2022, doi: 10.1007/s12652-022-03899-6.
[33] K. Haruna *et al.*, “Research paper recommender system based on public contextual metadata,” *Scientometrics*, vol. 125, no. 1, pp. 101–114, Oct. 2020, doi: 10.1007/s11192-020-03642-y.