Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 25, Issues 4, pp. 3181–3199, DOI 10.12694/scpe.v25i4.2882

AN AI-BASED CLASSIFICATION AND RECOMMENDATION SYSTEM FOR DIGITAL LIBRARIES

ABDULAZIZ I. ALOMRAN* AND IMTIAZ BASHA†

Abstract. The immense volume of online content linked to digital libraries has given emergence to the advancement of screening and recommendation systems. A recommendation system is vitally important in both academic institutions and e-libraries to assist professors, instructors, students, and researchers in finding appropriate sources of information. Distributed or collaborative screening is the most common method used in current recommendation systems. However, collaborative approaches cannot promote library repositories, including unrated or unpurchased electronic information. Thus, this paper deals with the automated classification and recommendation of a multiclass corpus found in virtual repositories (cloud databases). In various stages, Neuro-Fuzzy (NF) and Support Vector Machine (SVM) techniques are used as the base classifiers for the categorization of the essential subjects (contents). Later, a high-level ensemble learning strategy is utilized to recommend appropriate subjects from the available multiclass corpus. The methods use a CoC (Coherence of Content)-based inference mechanism to extract and filter the critical components before beginning the recommendation process. Experiments demonstrated that a recommended approach based on detailed conceptual descriptions instead of a handful of phrases/words might help academic and research communities to find relevant sources. Observing the results over a period of months shows that the suggested method increases user comfort, proving the system's acceptability to users in this way. In addition, compared to previous models, the accuracy in categorizing the requisite subjects is more than 97.16 per cent.

Key words: Digital Library, Accuracy, Cloud service, e-content, Webpage, Classification.

1. Introduction. Regular library services are migrating to the digital platform enabled with Internet service. All sorts of data are stored virtually in digital repositories that usually include user information such as profiles, reference histories, and document details, as well as users exploring and accessing log files and cataloguing information. The vast majority of the data is handled exclusively for user queries. Moreover, the usage of Web services to obtain information appears to be on the rapid rise. Consequently, digital libraries and Internet searches for library materials are becoming more popular.

Clearly, the Digital Library (DL) is a complicated domain in which a significant number of various disciplines, as well as specialties'', congregate with multiple attributes evident through its standard data structure. As a result of its interdisciplinary characteristics, the term "Digital Library" has evolved to symbolize diverse things to different people, each with its own unique viewpoint [1, 2]. The critical repercussion of this diversification is that, for the past decade, a significant number of DL systems were constructed in a sensible manner using specialized approaches that have been gained by integrating concepts drawn from many other sectors. It is remarkably complex to build DL systems that can be shared, reused, and developed collaboratively using a methodology that results in many disparate entities and unique processes for handling the required contents.

Most of the DL society has expressed the necessity to engage in DL system designs and create generalized DL management platforms with all the main elements that seem vital in enabling the complete scope of DL activity in many conceivable situations [3]. For this reason, DL management platforms should have features that address both general problems and the common elements used in several diverse issues. Aside from this consideration, its architectural design must be flexible enough to accommodate the specific needs of any given environment [4].

DL improves the capabilities of remote and virtual learning services [5]. The virtual learning approach illustrates that such infrastructure can be used to provide a diverse array of advancements to anyone's skill set.

*College of Humanities and Social Sciences, Department of Information Science, King Saud University, Kingdom of Saudi Arabia (benomran@ksu.edu.sa.)

†Library Director, Central Library, Mustaqbal University, Kingdom of Saudi Arabia (ibsahib-t@uom.edu.sa)

A wide range of virtual or e-learning materials from the DL system can be obtained, including e-documents via advanced web services, virtual communities, plus hypermedia. Most of the core components of DL are research/educational resources that can be shared across a variety of technological platforms. Formative assessment, corporations, federal agencies, entities, residences, and hubs of communal locations are all included in the domain of DL systems. DL could well be utilized to increase the efficacy of background and academic approaches to societal, technological, and research demands by instructors of host programmes [6].

Many researchers have identified the following advantages of DL based on the e-Learning paradigm [7]:

- In-depth training in innovative and collaborative learning with a continuous enhancement for participants.
- Enhanced capacity for information acquisition, expression, and comprehension.
- Learners' long-term drive to learn about complexity and collaboration, as well as their ability to examine critically and creatively.

On the other side, DL users can benefit from specific suggestions (recommendations) to retrieve required documents from a vast repository. In a conventional DL system, a user could search for a particular term and get outcomes based only on the terms entered. In this case, the user would have a hard time finding essential documents that weren't related to the search queries. The document recommendation service provides the user with a variety of possible contents of DL that are centred based on their own interests, lending records, or the identities of other users having similar interests. This enables users to find interesting documents they may not have previously fed as input terms. Since current DL users are likely to lend more significant numbers of documents, the new users will likely utilize the platform; the library administration can expect an increase in document checkouts. The amount of documents circulated in the larger society and the advancement of knowledge has made it more difficult for the DL to arrange and operationalize the soft contents. For this reason, it is vital to analyze the user's identities (profile) and associated metadata (queries applied, content demand, and so on) in order to predict their interests. Thus, such kinds of predictions and recommendations are possible using Smart Recommendation Systems (SRS).

1.1. Need of SRS for Digital Libraries. Smart Recommendation Systems (SRS) play a crucial role in various aspects of DL systems. They cater to the needs of both customers and businesses by providing personalized recommendations. These systems collaborate with diverse archives and repositories, offering users valuable suggestions based on their past preferences and sentiments. With the rapid growth of the internet and platforms like Twitter, Amazon, and Facebook, users now have a plethora of options to share their expertise, facts, and opinions. SRS helps manage data overload and facilitates multi-user relationship governance. It is widely beneficial in sectors such as e-library, e-regulators, e-business, and e-multimedia, aiding in cataloging management and knowledge representation techniques.

According to [8], an SRS may assist users in defining their interests, locating appropriate content, and aiding in interactive study activities. In current history, RS has established itself as a unique instrument that may be used to address the issue of data overload. When it comes to solving such problems, the role of recommendation systems is crucial [9]. RS may help users discover new, previously unknown content that is relevant to their present work. Users benefit from using RS because they assist them in locating required content that is not available elsewhere. Yet, many crawlers use RS to filter and locate non-conventional content. In 1990, JussiKarlgren, a professor at the University of Columbia, introduced a "Digital Bookshelf", in 1990 was the first reference to (RS) [10]. Subsequently, a number of notable specialists have added their insights. A significant amount of technical progress has been made during the RS conceptual transition. The recommendations that are made via DL systems must abide by the users' fundamental requirements [11].

- Users' prior knowledge.
- The system incorporates conceptual ideas.
- Information is needed to build a framework of conceptual understanding.

1.2. Categories of SRS. There exist a few significant categories of SRS; they are collaborative filtering, content-based filtering, and hybrid filtering. Therefore, gaining the base knowledge of all the filtering RS is essential. The classification of different filtering RS is depicted in Figure 1.1.

1.3. Content-Based Filtering RS. Contentcentric Recommendation Systems (RS) analyze the client's interests and offer tailored recommendations based on individual preferences. These RS store data from pre-

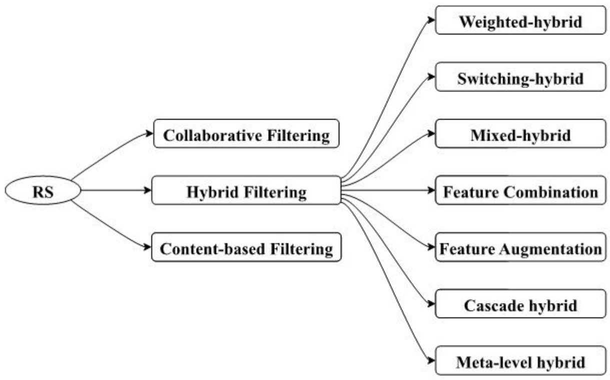

Fig. 1.1: Various RS Filtering Techniques

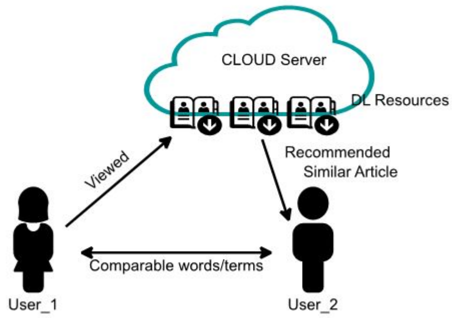

Fig. 1.2: Content-based Filtering RS

vious tasks in the user's record, providing a diverse range of entities and resources. Content-based filtering is extensively used to filter recommendations by focusing on the user's past records and the core content of documents shown in Figure 1.2. The dataset, consisting of phrases or words forming a narrative, is created by identifying comparable words/phrases and splitting the viewed content. The process of substance-based segregation involves identifying words, distinguishing them from referred items, and establishing a hierarchy of intellects to consider user logs and provide personalized recommendations. Content analyses are often used as the primary data source, and term breakdown from user reports is a common procedure. Phrases represent incident reports as feature variables in a multi-dimensional space using word vector representation and passive semantic sorting.

A user's logs or records can be incorporated via a range of techniques, including significance critique, inherited computations, neural networks, and probabilistic classifications. In addition, record-based communication methods can use word vector representation and passive semantic sorting.

1.4. Collaborative Filtering RS. When a method falls under this category, it seeks to align a user's natural tendencies toward the interesting content with prior user actions. A vast proportion of user interactions, assessments, reviews as visitants and other forms of activity are analyzed before a consolidated server is delivered. At this step, it looks to see whether the user's perception is similar to that of another similar kind of user or neighbor. The basic premise of collaborative filtering is that a user could choose a privatized document's contents if another observer enjoys it just as much as they do. The recommender efficiency in a deriving technique may therefore be described as "people like you also viewed/referenced similar content" as a frame of comparison. Such systems are now widely utilized and have been shown to be quite effective for exploratory

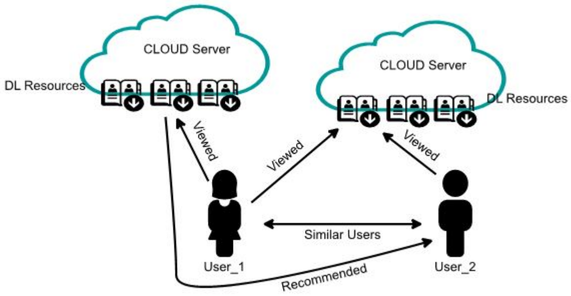

Fig. 1.3: Collaborative Filtering RS

searching and hence have risen in the marketing rating. To perform well, the platform needs an enormous amount of user-generated information-rich paradigms with the motivation of serving specific ideas. Because of the shrinking quantity of data, collaborative filtering is unable to discover the required content that nobody has looked at. In order to reduce the amount of frosty data points, various RS incorporate default procedures into the user experience, for example, by integrating them as the main page or adapting them into base content-building. While collaborative filtering can produce fresh, high-quality ideas, it needs some user information to recommend suitable content. It's much more challenging to use it on websites that don't need a user to log in. Figure 1.3 depicts the concept of the collaborative filtering technique.

1.5. Hybrid RS. Hybrid RS is a strategy that combines the best features of many different RS in order to meet the distinct requirements and concerns of a specific user. In particular, it provides numerical and statistical processes to demonstrate the optimal recommendations for user requirements. A mixed or hybridized RS tool that integrates a minimum of two recommendation methods is more likely to perform well than the other RS that relies on just one. Most of the time, neighborhood filtering data is acquired in conjunction with another technique in an effort to avoid the escalation of other problems.

Weighted-hybrid RS: Each model's outcomes are combined to provide a weighted RS that does not alter in weight between the training and testing sets. One benefit of using the weighted RS is that it lets users combine different approaches to help the available facts in a way that is additive (linear), especially during the recommendation phase.

Switching hybrid RS: Depending on the scenario, the switching hybridization uses a single RS. Users must configure the decision support eligibility requirements as per available patterns or other information to build the model for the component-level sensitive database. This strategy adds a new layer to the RS, which chooses the best model to use.

Mixed hybrid RS: A mixed-hybrid strategy begins by generating a range of potential datasets based on the user's portfolio and characteristics. Consequently, the recommendation engine receives various possible inputs, which are then used to integrate the predictions to arrive at a final suggestion. This RS model is capable of producing a high series of recommendations concurrently and sometimes adapts to the fragmentary information with a suitable model to improve its performance.

Feature Combination hybrid RS: In a feature combination hybrid, the user adds a virtualized contributory prediction algorithm to the recommendation framework to perform as a component synthesizer concerning the initial user record set. For instance, collaborative features can be added to a content-based approach. Furthermore, this model can take into account the information from the suggestive module, so it doesn't have to depend on just one model.

Feature Augmentation hybrid RS: A feature augmentation hybrid strategy is utilized to establish a ranking or categorization of the client portfolio that is subsequently incorporated into the primary RS to predict the overall result. No significant changes are made to the basic recommender systems when using a feature augmentation hybrid, yet it still enhances the system's predictive performance.

Cascade hybrid RS: In the case of cascade hybrid RS, the principal RS is responsible for delivering the primary outcome. Later, the system is allowed to utilize the auxiliary model to address trivial concerns, like dissolving a deadlock in the score. Thus, the model exhibits a hierarchical-based operational mechanism. Also, when most datasets are small, the auxiliary model can help deal with problems like not enough data or scores that are almost the same.

Meta-Level hybrid RS: This type of hybrid RS bears some resemblance with the functional procedures of feature augmentation in the sense that the contributory engine supplies the primary recommender system with an enhanced dataset. But unlike feature augmentation, the input feed to the primary recommender model is replaced by an alternative model, which is seen as a learned contributor model.

However, any method can only provide recommendations for the viewer depending on the preferences that they already have. That is to say, the system is unable to go beyond the preferences already expressed by the users. As a logical consequence, the intention of this study was to examine the most emerging trends in machine learning (ML) due to their high suitability for the content recommendation processes in DL systems. The area of ML is most rapidly expanding in the computing industry. They are a component of artificial intelligence that was developed to offer systems some characteristics of human intellect. They build analytical models to make inferences and forecast results based on testing facts that they learned. Because of their precision and effectiveness, they have been used in a wide range of data processing and predictive activities. Such new technologies have enhanced recommenders' prediction and accuracy.

In this work, an ensemble learning (ESL) technique is used to identify suitable content areas from a multi-labelled archive, which is organized into two stages. The Support Vector Machine (SVM) and Neuro-Fuzzy (NF) approaches provide the basis for the initial stage (first-level) of ensembling. In contrast, the stacking ensembling approach is the central part of the second stage. Extracting and filtering all relevant data from a repository is a necessary step before the commencement of recommendation procedures. Thus, for that purpose, we utilized CoC, which performs conceptual filtering rather than focusing on individual keywords. It also works with the rationale and suitable categorization of queries/requests and relevant content (all associated digital content) to provide valuable and intelligible information.

1.6. Research Motivation and Contribution. In digital learning systems, learner preferences and habits are considered when delivering services. However, the focus on individual abilities is often overlooked in recommendation methodologies. Some learning techniques overload users with excessive linkage patterns, creating information burdens. The main research concern in online learning is the lack of a proper Recommendation System (RS). RS can provide recommendations based on user cooperation and offer some control over preferences, but there is also a need for regulation. This work proposes a smart recommendation system (SRS) that considers user abilities and preferences for online and offline complexities of digital learning content. It introduces an ESL technique to recommend suitable difficulty levels based on user traits from a large pool of data repositories. Observations suggest that this recommended RS has the potential to yield excellent results, facilitating efficient information retrieval for users and learners. The prominent objectives of the research work are,

- To increase the precision and dependability of content suggestions for readers and researchers using Ensemble Learning Strategies (ESL) to boost the performance of recommender systems (RS) in digital library settings.
- To enhance and verify the efficiency of the ESL-based digital library system, we will conduct thorough testing using a variety of datasets. Our goal is to minimize error rates, such as fallout, and achieve high accuracy. This will ultimately enhance user confidence and streamline accessing pertinent data from a wide range of digital content.

The entire article is organized in such a way as to make the reader understand the core concept of the work and to lead them to the future research process. Section 2 delineates the relevant work that has been carried out in recent times and identifies the research gap. Section 3 outlines the preliminary knowledge required to implement the proposed model. Section 4 elaborates on the core mechanism of the proposed methodology with sufficient information. Section 5 discussed the impact of model performance regardless of the objectives and explored the observation to the maximum degree. Section 6 summarizes the endnotes of the research work with possible future work.

2. Relevant Studies. As suggested [12] a paradigm for recommending individualized learning materials to students based on the specification and credibility of information. An investigation with the model indicated that obtaining resources that were appraised favorably increased the learner's contentment when esteem information was used. [13] also developed a collaborative filtering-based item RS. Search engines and content processing were reduced in complexity by using LDAP (Lightweight Directory Access Protocol) and JAXB (Java Architecture for XML Binding) approaches.

Using genetic computations [14] and a few additional suggested strategies, the preliminary model outperforms the previous techniques in obtaining data via Google Maps. Furthermore, with the use of bio-inspired grouping approaches, experts have been able to discover the best possible suggestions for a given grouping system. Different methods have been used in the past to solve the grouping problem, but now bio-inspired procedure requirements are being set up to make better recommendations.

To generate suggestions, investigators [15] used a technique focused on client grouping to examine "m" individuals as well as "n" objects. A K-means categorization strategy is also used to classify the visitors based on their preferences. Finally, it creates a suggestion using a brand-new technique called the voting process. A novel collaborative screening method is tested against a conventional approach. The findings have shown that the method is not only faster but also produces an outcome with higher precision than it was before. In this case, the phasing of the content is critical to creating more effective and tailored electronic learning systems. As compared to a traditional, unsorted series of learning resources, collaborative filtering that is more successful will provide recommendations for an instructional process that is referred to as the study route. This learning series needs to be in an appropriate sequence and include both a beginning spot and a concluding spot. Moreover, a proposed sequence must be tailored to the study interests of the participants, in particular, to boost their competence to gain knowledge. In addition, the duration of this sequencing is not cast in concrete for all participants since individuals vary in variables such as their degree of expertise, the preferred method of studying, and emotional state at any required time.

For e-learning via digital platforms, researchers [16] utilized a variety of strategies that have been examined to implement the recommendations that seem to be tag-based. Depending on vector (tensor) discretization and a refined method, a model ordering (ranking) was devised for generating efficient suggestions. To reduce the processing time as well as resource utilization while ensuring the integrity of the recommendations, this research work also advocated shrinking the tag range and using a grouping approach that relies on a preferred learning concept.

Introduced a novel [17] varying length optimization approach that has been employed to represent the Instructional Route RS that is utilized for digital learners. This representation considers the varying learning patterns as well as the different degrees of expertise. The research's outcomes confirmed the better efficacy of Instructional Route RS in a digital learning context. The RS is useful in improving user engagement. Only recently, the strategies adopted by [18] have suggested the Bat Computation methodology, which is intended to estimate the scores of the objects (or attributes) to determine a preferable neighborhood for proactive users. In conclusion, it was found that the intuitive techniques offered scores for numerous factors in order to produce tailored suggestions. By comparing the framework's functionality with the existing methods, it is proved that the BA scored 6.9 percent higher on mean absolute margin and an F1-measure. Furthermore, user interests in various contexts are taken into account by contextually sensitive RS. Whenever these methods were being designed, the absolute priority was to preserve the credibility of recommendation engines and drastically reduce the intrusions of prejudiced users that impacted the mechanism and its conclusion. Additionally, the core target was to reduce the number of skewed users. Besides this method, techniques exist for detecting malevolent users in various situations and contexts.

The research work of [19] exhibits the k-means grouping and bio-inspired strategies, which were used in a novel hybrid approach for a unique hybridized recommender method tested on the sample set of Movie lenses. This new deterministic approach obtained good results in the aspects of scalability and versatility. It made tailored film recommendations and made the cold-start problem easier to deal with.

An efficient method for selecting essential features has been devised by the authors of this research work [20]. This method is used in two different versions centred upon Artificial Bee Colony (ABC) employing Genetic Algorithm (GA) as well as (particle swarm optimization) PSO, respectively. These models are referred to as the ABC-GA and ABC-PSO variants. Improvement procedures of the ABC approach have been included to achieve a perfect blend between extraction and investigation. Because of this, a subset of features was deemed to be effective in regard to precision throughout the experiment. Furthermore, web 2.0 content can now be accessed and organized using a new mechanism known as tag grouping. Therefore, this methodology could be used for efficient social labelling/tagging activities.

Implemented [21] an artificial intelligence-powered open-source conversational software platform named Rasa. The researcher advised its potential adoption by libraries as the author introduced the core of chatbot technology to apply to day-to-day / daily library applications. Further, the researcher concluded that AI-based conversational software have need of more profound investigation in order to perfectly study and emulate human conversations; it, however, has limitless controls to prompt action-packed services and to function in a way to satisfy the patrons of the library.

Introduced [22] a new working mechanism where a competency assessment was used to verify the students' behavior and academic abilities, and the assessment results were used to develop study materials. Tailored Ranking (page) and the Adaptive Knowledge Scheme, both of which use real-time personalization, provide the backbone of this approach. In the event, the Navy's Bayes predictor was used to categorize the students based on their performance on the competence exam. High-skilled users are given cutting-edge materials, whereas intermediate and novice users get less advanced materials. All of these factors are considered when evaluating how the students performed.

Collaborative Filtering and Content-Based Filtering, as well as Hybrid approaches in recommendations, are the commonest. The Content-Based technique relies on the user's plan as a guide. The previous operations, such as user rating, are practiced by Content-Based-based systems without the inclusion of user input. Hybrid approaches combine Collaborative and Content-Based Filtering procedures to achieve the best outcome [23].

In online learning [24], a Hybrid Action-Related K-Nearest Neighbor Similarity (HAR-KNN) recommender was suggested, which centralizes the efficiency of hybridization processing to enhance user activity lattice by generating the matrix of attributes. Race classifications will be used to sort the characteristics according to their quantity as well as performance. The suggested technique also overcomes the issues of the prior approaches in assessing user preference on items and balancing attribute evaluation effectively. When it comes to determining user usage statistics that correlate to certain user groups, the actual and digital content using the K-NN classifier has been shown to be an effective tool. Outliers and prediction metrics are used to assess the intended empirical outcome.

Successfully [25] extended the intellectual notion of RS to create a revolutionary SRS. Recommendation procedures are selected based on a hybrid strategy that incorporates and handles all critical data and individual recommendation requirements. Combining interactive filtering, information, and expert techniques is feasible with the hybrid version. Using four components, the data is filtered: the scenario, learners, curriculum, and digital materials; determining factors like socioeconomic factors; link traits; locale; and individualized learning preferences. Trials were performed to develop a curriculum to examine the architecture's intellectual ability and autonomy.

Have focused their [26] research on expertise learning trajectory recommendation strategies that do not produce multimodal instructional routes to meet various learning requirements in reality (in the real world). For this reason, a paradigm that recommends studying paths relying on multivariate domain knowledge is being developed. This approach maintains knowledge items independently and organizes them into several groups. The information graph then suggests six significant conceptual relationships among various learning items. Another method is to use multivariate knowledge base architecture to develop and offer tailored learning pathways depending on the e-target learner's training objective. This approach is referred to as the test pattern recommendation approach. The results of the research show that a system based on the data collected during the experiments can make and suggest good learning paths for online learners that are both competent and relevant.

As per the research work of [27], an iterative approach with content-based technique and vector decomposition features is used in a blended system to improve the resilience of cooperation filtering. Furthermore, the authors develop an incremental assessment technique for actual statistics while researching to test the outcomes. The report's hypothesis outcomes indicate that the highlighted blended system can be developed as a potential

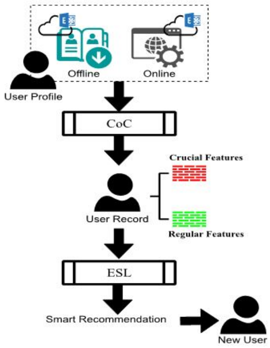

Fig. 3.1: RS Process Flow

research route. The presented method achieves superior outcomes in terms of both gaining knowledge rate and effectiveness.

Multi-label k-nearest neighbor (ML-KNN) and classifier chain k-nearest neighbor (CC-KNN), combined with latent Dirichlet allocation (LDA) implemented in order to avoid manual classification [28]. The researchers pointed out that manual classification is a time-consumption and expensive as the number of patent documents is increasing day by day. Further, researchers specified strongly that automated classification is indispensable to avoid such complicated assignments. According to the authors, the Automated classification harvests precise outcomes and based on it; they implemented quoted the above two algorithms.

Studies show that content phasing has a significant role in creating more effective and tailored RS. RS should propose a knowledge pattern through the preferred route, which must be in a reasonable order coupled with an appropriate beginning and finish point. The central objective of this research is to develop perfect RS for newcomers about how to use RS for reporting preferences in the existing DL models.

2.1. Summary of Research Review. Many RS do not allow for a tailored ecosystem to provide referrals to a population of learners focused on individual needs and their ability to encourage smart studying; this is a serious shortcoming. Giving learners/DL users relevant and reliable information is frequently a fundamental difficulty in DL systems. Huge data computation and code generation are required to handle complexity, training rate, and integration of enormous amounts of information with proper correlation to satisfying the demands. Any suggested system should proactively recommend substantial supplies to the user to increase the usability of DL. Personalization of RS via reactive training layout and tailored content are two of the most exciting aspects of any upcoming approach.

3. Methodology. The phrase "Digital Library-DL" is commonly alluded to as "e-library", attributed to the reason that digital content distribution and its acquisition or digital repositories are sublines of the online platform of digitized objects. Such objects include text, static visuals (images) in various formats, multimedia content, and digital data accessible over the cloud services. The advancements in cloud computing have breathed fresh life into the process of creating digital libraries. Researchers have developed an RS for digital libraries that considers the unique needs of each user and actively sends the most relevant information based on their preferences. In order to achieve this, the library's volumes, techniques, search documents, periodical documents, and other data repositories must be combined with the user's patterns and preferences. In addition, users' personal information and related procedures can be stored on the cloud infrastructure.

Usage records are employed to build user records that are maintained in either cloud-based virtual repositories or on-premises. Log files are used to store actual facts and essential data on visitants' activities in DL, including both static data and dynamic web-oriented information. In addition, retrieving the necessary information takes time and constitutes a complicated procedure for DL users. As a practical issue, the recommendation engine forecasts the user's interests and recommends the most relevant material as a result. In this regard, creating a user record, collecting data needed to develop a suitable RS, and retrieving and filtering

Table 3.1: Key Features of the Digital Library Datasets

	Features	Offline Log Files	Online Log Files
Regular Features	User Components (user ID, catalogue numbers)	$U = \{a_1, a_2, \cdots a_n\}$	
	Average page time	$\mu_{p_t}\left(f_p\right)$	$\mu_{p_t}\left(f_p\right)$
	Average Site time, $S = \{w_1, w_2, ., w_n\}$	-	$\mu_{s_t}\left(O_s\right)$
	Average visitant counts/Preference Count	$\mu_{V_t}\left(f_V\right)^{p_i}$	$\mu_{V_t}\left(o_V\right)^{p_i}$
	Exit Rate, E_r	-	$\sum_{U=1} l/\tau$
Crucial Features	Average page rank	-	$\mu_{R_t}\left(O_R\right)$
	Frequent content terms accessed at P_i	$\sum_{i=0}\left(F_t\right)^{f_p}$	$\sum_{i=0}\left(F_t\right)^{O_p}$
	Mean similarity among various content terms	$\mu_{\delta(c_1, c_2, \cdots, c_n)}\left[f_p\right]$	$\mu_{\delta(c_1, c_2, \cdots, c_n)}\left[O_p\right]$
	Mean of Frequent term utilized among various users	$\sum_{U=1}^{N} \mu\left(F_t^{f_p}\right)$	$\sum_{U=1}^{N} \mu\left(F_t^{O_p}\right)$
	Bounce Rate, B_r	-	$\sum_{U=1} \xi/\tau$
Total Instances		8134	1562
Total Instances per classes		10%	10%
*'a' denotes components, 'f' and 'O' denote offline and online log files			

pertinent data are always perceived as crucial. A CoC inference system is used to extract and select the required content in this model. The suggested RS process flow is shown in figure 3.1.

Three content-rich datasets are incorporated in this work to train the proposed RS model. The datasets are obtained from [29]; Cuneiform Digital Library Initiative, (n.d.) [30]; and Book Recommendation Dataset, (n.d.) [31]. All three datasets contain the required digital library component, comprising metadata, multidisciplinary literature under various catalogues, content ratings, user profiles, etc. As an initial process in the proposed RS, the appropriate information is extracted and filtered from the considered datasets using the CoC inference system. Then, CoC utilizes logs/records files from the datasets to filter the required content. Finally, the log/record files are created based on the access history, user's interests, and standard access phenomena in a unique database.

Digital library log/record files are stored and accessed via a cloud platform which is enabled with internet service. Thus, the files are viewed in two aspects: offline and online log/record files. The offline files are stored files, whereas online files are dynamic files accessed via the internet, and the associated log/record files are periodically updated. During the extraction process, a few prominent features of both offline and online log files are considered. Those notable features are represented as regular and crucial features. Table 3.1 illustrates the key features and their description, which all influence the training process of the RS model.

Bounce rate estimates the number of users departing a webpage (T) without accessing it in any manner (cumulative one-page visit, ξ). Exit rates measure the proportion of visitants (T) who departed the webpage after seeing a specific page (l), regardless of the number of webpages they had previously seen in the particular session.

In the process of creating a user portfolio, a strength component (ψ) is allocated to each class of features. Selective traits are given more strength by adding additional values. This tends to boost the precision of RS. The proposed system employs a ψ value that ranges from 0.0 to 0.1. The configuration of ψ in the described approach varies based on the available features, which are stated as follows:

- ψ is specified as 0.1 for the most crucial features (significant elements) of a user record in an attempt to boost its impact during the recommendation process.
- ψ is specified as 0.5 for the regular features of a user's record. The act of boosting with this strength nominally powers its impact during the recommendation process.

3.1. CoC Inference Process. CoC is the degree to which terms within a phrase or sentence are coupled to form connotations. In this way, it's connected to the formal principle of coherence. The cohesiveness ratio is the totality of the relationships (similarities, δ) among all terms of content divided by the overall amount of possible terms of content. Considering a content that contains 'n' terms, the formal definition of the degree of correlation between two terms of that content, C_i and C_j, which refer to different sets of terms, T_i and T_j, respectively, is expressed as in equation (3.1),

$$\delta\left[(\mathbb{C}_i),(\mathbb{C}_j)\right] = \left\{\left|(T_i)\cup(T_j)\right|_{/\mathrm{n}}\right\}\cdot\left\{\left|(T_i)\cap(T_j)\right|/\min_{(}\left|T_i\right|,\left|T_j\right|\right)\right\} \tag{3.1}$$

The next level of the proposed procedure deals with the user's record that comprises preference factors and computations of CoC strength factors as input feed to the ESL model. As mentioned earlier in section 1, the proposed RS includes two ensemble processing stages, wherein the first stage employs the NF model and SVM techniques as the base classifier for recommending the preferred contents in the DL system. Later, utilizing the outcome of the first stage, the second stage, signifying ensemble technique, computes the available data and produces the optimal recommendation to the user, which is delineated elaborately in section 4.

In this section, a smart Recommendation System (RS) has been developed for digital libraries, considering user needs and preferences. User records, including personal information and browsing history, are stored either on-premises or in cloud-based repositories. The RS process involves creating user records, filtering data, and using a CoC inference system to extract content. Key features like page rank and frequent content terms are considered. The RS utilizes ensemble processing stages to recommend preferred content in the digital library system.

4. ESL Model. To improve the effectiveness of recommendation systems, techniques based on complementing the thematic information datasets are needed to be used. Thus, in this section, we elaborate on the core process of the ESL mechanism in recommending the appropriate content to the user in the DL platform. Out of two stages in ESL, the first stage comprises SVM and NF techniques for recommendations.

4.1. SVM in RS. Few studies have addressed the utilization of an optimized SVM concept for recommendation tasks. [32] presented the SVM, a revolutionary ML concept for learning digital content. In current history, the computing of SVM has evolved as one of the most powerful strategies for predicting and categorizing class labels. The SVM learns a splitting hyper-plane in order to optimize the error gaps and achieve high generalization. Many applications of the SVM have been practical recently, yielding impressive outcomes.

Kernel operations and computational processing are two critical components of SVM's deployment in RS. A non-linear separation plane in the high-dimensional feature space is created by employing kernel operations. Suppose the cost of computing the kernel functionality is comparable to that of computing the input space. In that case, a non-linear intervention of this kind does not result in an improvement, especially in the computation complexity. SVM makes use of a polynomial equation to deal with the input feed of digital content that can be expressed as in equation (4.1),

$$\rho\left[(\mathbf{a}\mathbb{C}_i),(b\mathbb{C}_j)\right] = \left[\left(\mathbf{a}\mathbf{C}_i^T,\mathbf{b}\mathbb{C}_j\right)+\boldsymbol{\beta}\right] \tag{4.1}$$

where kernel ρ denotes the inner component of feature vectors in the training datasets, aC_i and bC_j represent the input attributes of the concerned contents C_i and C_j, respectively. signifies the degree of polykernel. In the SVM training, both crucial and regular attributes are utilized to recommend the required content. The SVM uses a hyperplane that separates with a given range to the learning sets nearest data point to provide recommendations for various types of content demand. Such data points are referred to as a "support vector". Using optimal margin, the hyperplane divides a collection of highly desired content from a collection of disregarded stuff. Suppose if there are k number of content demands from different users, which can be paired as $(a_1, b_1), (a_2, b_2), , (a_n, b_n)$, and a corresponding vector is depicted as $a_i \in C_i, b_i \in -1, 1$.

SVM is tasked with learning the mapping of pairs $(a_i - > b_i)$. In the finished result, all of the suggested hyperplane-boundary vectors are included. It is characterized as the spacing between the higher dimensional space and the closest data point, especially in the linear condition. Figure 4.1 demonstrates the sample graphical representation of the SVM process in recommending the content to the user.

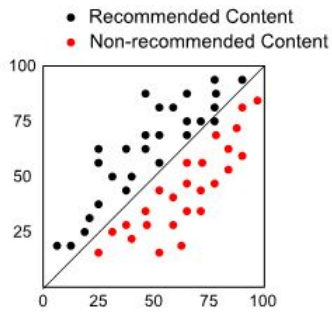

Fig. 4.1: Sample illustration of SVM in Content Recommendation for i^th User

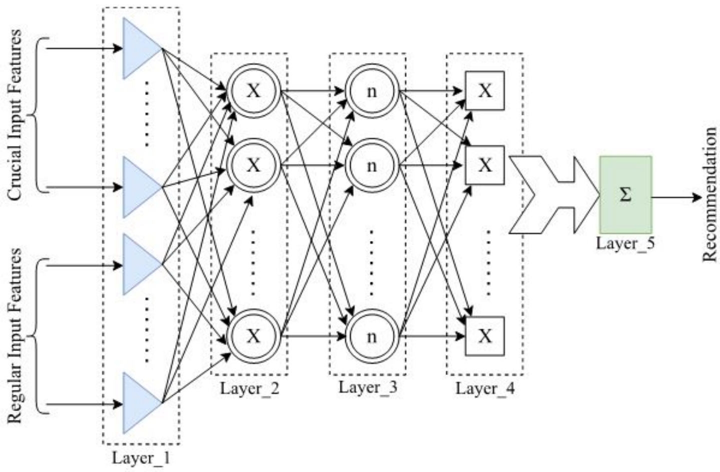

Fig. 4.2: ANFIS Architecture

4.2. NF. Adaptive Neuro-Fuzzy Inference System (ANFIS) was utilized as the second base classifier. All the computation processes of ANFIS were derived from the work of [33]. The membership function (mf), fuzzy inference repository, and rationale process form a significant part of ANFIS. In addition, ANFIS employs gradient descending to tune the vital parameters. To formalize the application of ML methods, the ANFIS design can be considered a multi-layered connectionist framework, as represented in figure 4.2. The model's knowledge base uses Takagi and Sugeno's [34] fuzziness models. Fuzzy sets correspond to both crucial and regular input feature variables, and fuzzy rules are made by linearly combining a constant and a data object. ANFIS follows a multi-layered (six layers) architecture mechanism that combines the basic computation of both fuzzy logic and neural networks.

Layer_0: A collection of input data that is fed to the ANFIS system is represented through this layer. Typically, this tier is not explicitly exposed in the primary ANFIS architecture.

Layer_1: Each incoming vector is measured as a fuzzy set, and furthermore, this defined mf is used to build this fuzzification layer. The mf accepts the input parameter and returns the input's degree of membership as an outcome ranging from 0 to 1. This result was obtained by fuzzifying the crisp data. Adaptive mf is applied to each unit at this stage and is expressed as in equation (4.2),

$$\mathrm{e}_{\mathrm{i}}^{l\mathbf{1}} = U_{[\boldsymbol{p}_j]}(\boldsymbol{x}) \tag{4.2}$$

where, $\mathrm{e}_{\mathrm{i}}^{l\mathbf{1}}$ indicate the node i's outcome, and $U_{[\boldsymbol{p}_j]}(\boldsymbol{x})$ denotes the mf's outcome of content preference set, P_j (fuzzy set) of users. This means that the component variables can be revised using gradients descendance instructional strategies for each mf which is continuous and piecewise distinct in characteristics. In ANFIS, several kinds of mf's are employed; however, simplified bell functions and Gaussian-based computation processes are frequently employed.

Layer_2: In the second layer, fuzzy rules are computed using the product of their respective outcomes from the preceding layer. The layer's i^th unit's outcome is shown in equation (4.3),

$$\mathrm{e}_i^{lz} = \left[\boldsymbol{U}_{[\boldsymbol{p}_i]}(\boldsymbol{x}) \cdot \boldsymbol{U}_{[Q_i]}(\boldsymbol{y})\right] \tag{4.3}$$

In equation (4.3), e_i^{lz} is the multiplicative result of the mf,$\boldsymbol{U}_{[\boldsymbol{p}_i]}(\boldsymbol{x})$ and $\boldsymbol{U}_{[Q_i]}(\boldsymbol{y})]$, which determines the intensity of each rule's output generation.Thus, for example, the output generation intensity of the seventh rule can be calculated as, $\left[\boldsymbol{U}_{[\boldsymbol{p}_i]}(\boldsymbol{x}) \cdot \boldsymbol{U}_{[Q_i]}(\boldsymbol{y})\right]$. There are several other fuzzy operators (e.g., min) available to attain this result.

Layer_3: Each fuzzy rule's normalized activation intensity is represented by every i^th node of this layer. Normalized activation intensity is computed by the jth rule's activation intensity divided by summing the output intensity for all formulated regulations (j^th rules). In this layer, in general, the j^th unit generates the following outputs,

$$\mathrm{e}_j^{l3} = [\phi_j / \Sigma_{i=1}^{r}(\phi_i)] \tag{4.4}$$

In equation (4.4), ϕ_i denotes the output triggering intensity of any i^th rule, r.

Layer_4: There are units/modes with updatable attributes related to this stack, making it adaptable like layer 2. Each node's result is a sequential transformation via linear function and is defined as follows,

$$\mathrm{e}_i^{l4} = \left[(\alpha_i x + b_i x + R_i) \times \bar{\phi}_i\right], \text{ where } \eta = (a_i x + b_i x + R_i) \tag{4.5}$$

In equation (4.5), represents the result of the i^th unit of this layer, whereas a_i, b_i, and R_i denotes the coefficients of fuzzy operation functions, which are fine-tuned at the attribute optimization processes.

Layer_5: As the total of all incoming signals, this layer contains one node that quantifies the aggregate outcome of this tier using the summation of all the signals generated in the previous layer, which can be expressed as ϕ,

$$\mathrm{e}_i^{i5} = \left[\sum_i [\boldsymbol{\eta} \times \phi_i] / \Sigma_i(\phi)\right] \tag{4.6}$$

In equation (4.6), the attainment of the final recommended outcome is denoted as e_i^{i5}.

4.3. ESL. ESL is a sophisticated ML approach that uses a unique blend of more than one base classifier to address a specific issue in computational learning. In ESL, there are indeed a variety of strategies to combine models, including boosting, stacking, and bagging. Stacking is a prominent and widespread ESL strategy [35] which effectively predicts the most complex outcome from heterogeneous data features. The final prediction is made via constructing a new model (meta-classifier) and enhancing the system's effectiveness. To boost performance, we can aggregate the results of ANFIS and SVM models that have been trained to address comparable problems. This technique is known as "stacking." The proposed stacking-based ensembling technique is illustrated in Figure 4.3.

For a better perspective, the fundamental steps of the generic stacking technique are shown via pseudocode in Table 4.1 In the initial phase of this process, the first stage base recommenders are modelled by utilizing the actual dataset. Then, SVM and ANFIS, the primary recommenders, undergo a training process for optimal recommendations. Finally, the weighting pattern of the input dataset is dynamically adjusted to match the primary predictors. Because of this, the algorithm's parameters are tuned in order to get more accurate results.

Compiling the outputs of base recommenders generates new datasets. Primitive categories from prior learners are used to build the subsequent level of recommenders. In contrast, estimates of the first stage recommenders are used as a newer feed to the second stage model. It is preferable to utilize recommenders'

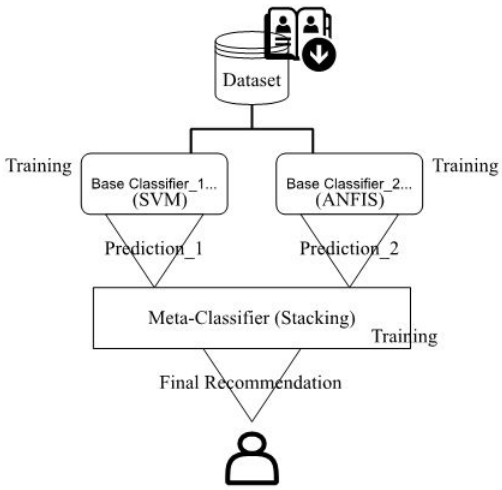

Fig. 4.3: Generalized Stacking Ensemble Technique

Table 4.1: Ensemble Learning

Input: $I_d = \sum_{k,l=1}^{m} (\delta_i) \in \gamma_i (\mathbb{C}_i, \mathbb{C}_i) \gamma_1 \leftarrow SVM; \gamma_2 \leftarrow$ ANFIS; Output: $\mathcal{G}_c$ //ensemble recommender
Step 1: $\forall : I_d - >$ (1 to n)do Step 2: $\zeta \rightarrow [\gamma_i \mid d]$; / learning phase Step 3:$\forall n \rightarrow (1 \text{ to } N)$ do Step 4: $Z_d \leftarrow \left[\left(N_n^2, \varepsilon_l\right) \mid d\right]$ // new dataset Step 5: $\boldsymbol{G_c} \leftarrow \left[\left(d_j^{\text{new}}\right) \mid \left\{N_n^2, \varepsilon_l\right\}\right]$ // Meta-classifier Step 6: Return G_c; / Final Recommendation

likelihood estimates rather than projected labels. The non-linear activating factor of softmax was used to create additional attributes. Next, a more effective targeted class is forecasted using a newly obtained data set.

In summary, The ESL mechanism combines Support Vector Machines (SVM) and Adaptive Neuro-Fuzzy Inference System (ANFIS) to enhance recommendation systems in digital libraries. SVM uses a splitting hyperplane to optimize error gaps and recommend preferred content based on user demands. ANFIS employs fuzzy logic and neural networks to make recommendations using fuzzy sets and rules. Through stacking, the outputs of SVM and ANFIS are combined to create a meta-classifier for more accurate predictions. This approach improves recommendation effectiveness in digital libraries shown in Figure 4.2.

5. Performance Measures and Analysis. The proposed model was created for SRS using ML to enhance the content recommendations in DL. This section's analytical review serves to demonstrate the validity of the work. Using distinct data sets, the findings of the study are obtained. To illustrate the efficiency of the suggested model, the parameters of the datasets were categorized based on two attributable factors: crucial and regular attributes in each trial. In evaluating the proposed approach for diverse performance metrics, training, as well as test sets from datasets, is combined in various ways with distinct combinations. However, in each research phase, the weights attributed to user records and content parameters are maintained constant since preferable weight configurations have previously been tested by determining the semantic similarity among individuals and various content. Assessments of attained precision and errors [36] are used to verify the actual facts of these investigations. The suggested model's outcomes are ultimately compared to recently reported, state-of-the-art approaches. The models compared to the proposed ESL model are the standalone performances of SVM, ANFIS, HAR-KNN, and K-means. Metrics like accuracy, Mean Absolute Error (MAE), and Root

Table 4.2: Specific Requirement of the Study

Aspect	Recommendation
Tool	TensorFlow or PuTorch
Libraries/Packages	scikit-learn, NumPy, pandas, matplotlib, fuzzywuzzy
Programming Language	Python
Implementation	Develop a custom implementation using the chosen framework and libraries/packages
Parameterization	Experiment with different hyperparameters for SVMand ANFIS models (e.g., kernel type, regularization parameter, degree of polvkernel, number of layers and nodes in ANFIS)
Evaluation	Use appropriate evaluation metrics such as accuracy, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE), rate of fallout
Replicability/Reproducibility	Provide detailed documentation, code comments, and instructions to ensure replicability and reproducibility of the study

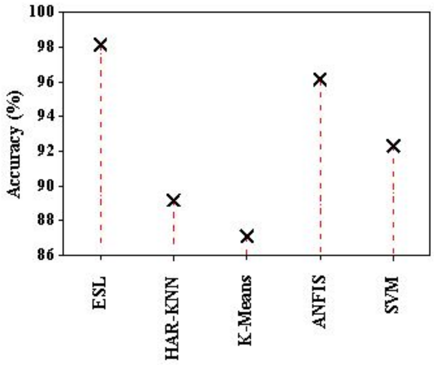

Fig. 5.1: Accuracy Analysis

Mean Square Error (RMSE) [37] measurements are used to assess the effectiveness of the proposed RS in the DL context. In addition, the fallout rate is included for analysis purposes. These metrics are referred to as efficiency checkers, and a lower error margin indicates that the approach is performing better overall. The in-depth analysis of this measure's outcomes is outlined in subsequent sections.

After several training and testing trials at each stage of the ESL models, comparable accuracy-based outcomes are obtained and shown in Figure 5.1. According to the result, ESL-based RS in the DL context obtained an optimal accuracy of 98.15%, which is 7 per cent greater than the prevailing RS. Also, as seen by the vertical drop line, alternative approaches have an average accuracy rate of 91.19%. Conventional techniques could not provide a precise recommendation, negatively impacting the platform's stability and efficacy in real-world scenarios. In ESL, users' records can be sorted by term similarity to determine which terms they commonly seek that are connected to one another. While searching for terms with numerous meanings in the lexicon dictionary, this function helps to prevent term-level uncertainty. Moreover, selecting appropriate features with fixed weightage aided the models in re-tuning the feature parameters to attain the optimal outcome (recommendation).

Figure 5.2 depicts the confusion matrix of RS, where it is observed that the results of ESL signify maximum coverage of topics in various fields and attained an average accuracy of more than 98%. Only the sociology and literature sector reports around 97% accuracy level which is due to the strong dependence of sociology on literature as a critical source of information.Initially, to investigate the model's potentiality in recommending the user's required library content, we opt to set four different training and testing proportions, (80:20), (70:30), (60:40), and (50:50), respectively. From Figure 5.3, when we analyze the performance of the ESL model using

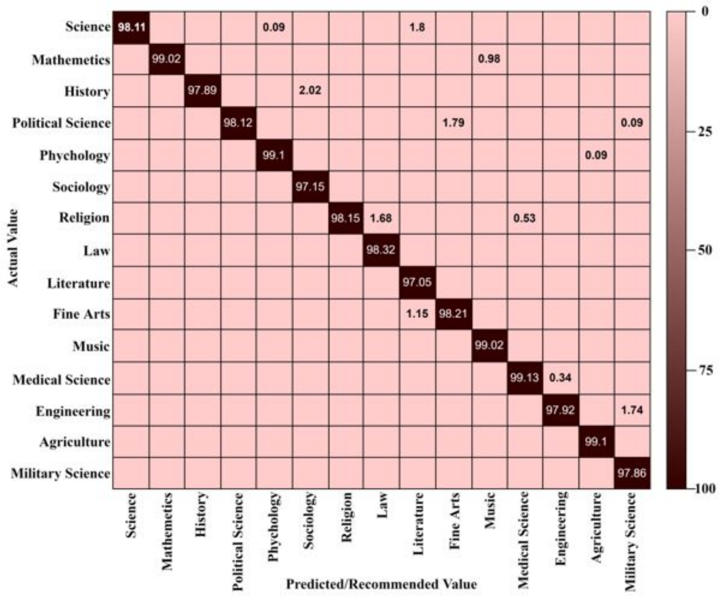

Fig. 5.2: Confusion Matrix

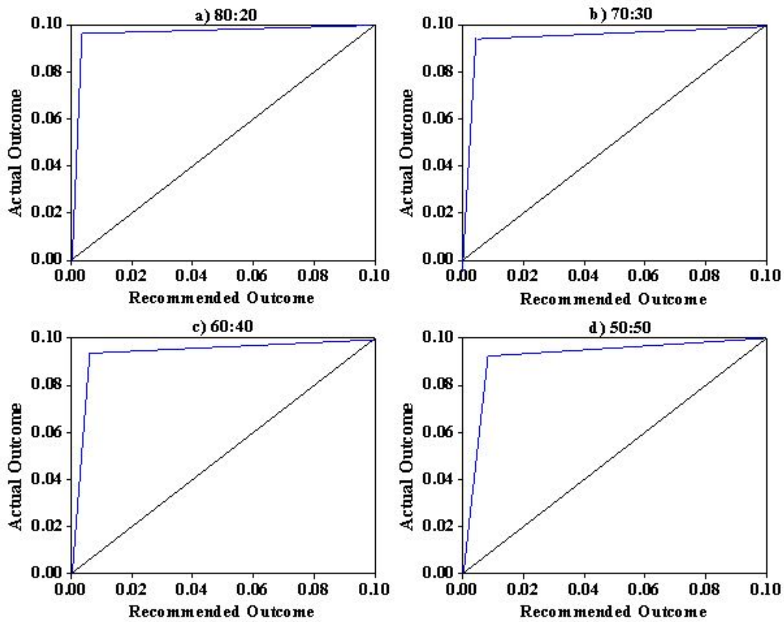

Fig. 5.3: Analysis of MAE at varying Training: Testing sample ratio

MAE, it's been noted that the model's performance is batterer at proportion (80:20). The model's efficiency degrades gradually as the ratio of the training dataset gradually decreases, which are evident from the outcome (Figure 5.3 a, 5.3 b, 5.3 c, 5.3 d). Despite the performance degradations, the model sustains to get least error margin (below 0.1).

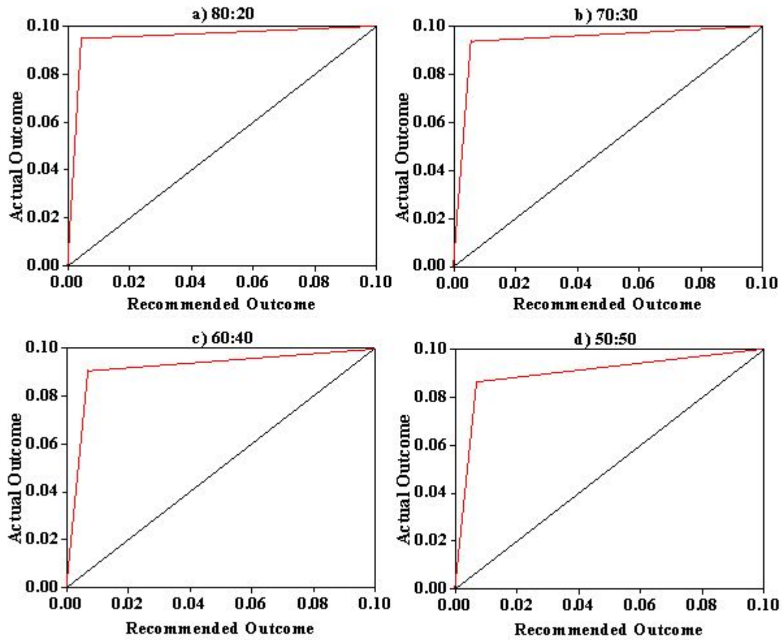

Fig. 5.4: Analysis of RMSE at varying Training: Testing sample ratio

The RMSE calculates the discrepancy between the actual results and those estimated. The variance of the estimation error or latent variables is computed using the RMSE formula. Researchers look at the accumulated depreciation to determine how far the trend lines deviate from the standardized residuals. Figure 5.4 shows that the RMSE improves with fewer latent irrelevant vectors (Root Mean Squared Error). However, it was also discovered that the RMSE error varies depending on the training-to-testing ratio, contrary to what would be expected. Though the minuscule error margins are maintained for all four proportions, it is necessary to note that the proposed model performs effectively when the ratio of the dataset is 80:20, which signifies that the model with a specific dataset proportion for training and testing has a mild impact on the intended outcome. But at the same time, if the sampling and learning rates are tuned as per the volume of available datasets, this kind of proportion's influence may deeply impact the desired outcome. From the resultants exhibited, it is noted that the error value associated with this metric has the most negligible significance to the RS since the reported value is below 0.09 for all the proportions; this indirectly signifies the precision level of the proposed model. The RMSE variance for a two-feature vector (crucial and regular) is lower, as seen from the Figure 5.4, (5.4a, 5.4b, 5.4c, 5.4d). For this reason, we decided to limit the range of subliminal variables to two; this is indeed a decent strategy because the range of implicit vectors selected must always need to be minimal.

Rate of Fallout (RF): It is defined as the recommended percentage of webpage's/e-content that was not pertinent to the overall count of the content/pages that were suggested. equation (5.1) is used to compute RF.

$$RF = \text{ falsePositive } /([\text{ truePositive }] + [\text{ falsePositive }]) \tag{5.1}$$

Figure 5.5 represents the outcome of RF, which insists that the proposed ESL maintains the minimized fallout (2.16), which shows more than 90% improvement compared to the other four models. All the other existing models exhibit average fallout of 6.43, which may considerably deviate from the outcome of the recommendation process. As a consequence, RF increases for any user requirements.

6. Conclusion and Future Work. It is becoming more important for researchers to employ recommender systems to expedite quality research and boost user confidence in finding the right information. Ensemble learning methods were employed in this work to boost RS's effectiveness. For digital library infrastructure,

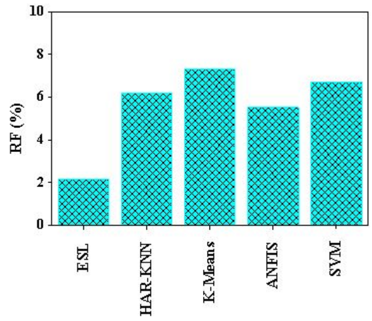

Fig. 5.5: RF Analysis

ESL-based RS was suggested in this research. The key strategies for optimizing the efficiency of an ESL-based DL system were also discussed in this paper. First, essential components are extracted and then filtered using a CoC (Coherence of Content) inference methodology. Next, we tested the conceptual model using three different datasets [29, 30, 31]. They were then compared to the standard method. The outcome of the random forest (RF) model, indicating that the proposed ESL achieves a significantly minimized fallout (2.16) compared to other models with an average fallout of 6.43. This highlights the effectiveness of ESL in providing accurate recommendations. It was noted that the ESL model performs best at the 80:20 proportion and gradually degrades as the ratio decreases and registers an average accuracy of over 98%. As shown by these findings, a proposed methodology may be used to provide suggestions for billions of pieces of content available across billions of documents.

Limitation: Although the study compares the proposed system with a standard method and utilizes multiple datasets, there is no mention of external validation against independent datasets or real-world scenarios. External validation is crucial to assess the system's performance in diverse environments and ensure its effectiveness beyond the specific datasets used in the study.

Future work: Re-tuning hyperparameters (such as learning rate) on our suggested model will be investigated in a future study. We would also like to use this idea to make document categorization frameworks available as a cloud service.

REFERENCES

[1] Kumar, V. V., Raghunath, K. M. K., Muthukumaran, V., Joseph, R. B., Beschi, I. S., & Uday, A. K. (2021). Aspect based sentiment analysis and smart classification in uncertain feedback pool. International Journal of System Assurance Engineering and Management, 13(S1), 252–262. https://doi.org/10.1007/s13198-021-01379-2

[2] Gonçalves, M. A., Moreira, B. L., Fox, E. A., and Watson, L. T. (2007). "What is a good digital library?" – A quality model for digital libraries. Information Processing and Management, 43(5), 1416–1437. https://doi.org/10.1016/j.ipm.2006.11.010

[3] Ioannidis, Y. (2005). Digital libraries at a crossroads. International Journal on Digital Libraries, 5(4), 255–265. https://doi.org/10.1007/s00799-004-0098-4

[4] Karthikeyan, T., Sekaran, K., Ranjith, D., &Balajee, J. M. (2019). Personalized content extraction and text classification using effective web scraping techniques. International Journal of Web Portals (IJWP), 11(2), 41-52.

[5] Pourjahanshahi, F., Mollahosseini, A., and Dehyadegari, S. (2023). Website quality and users' intention to use digital libraries: Examining users' attitudes, online co-creation experiences, and eWOM. Journal of Retailing and Consumer Services, 74, 103393. https://doi.org/10.1016/j.jretconser.2023.103393

[6] Abel, F., Bittencourt, I. I., iCosta, E., Henze, N., Krause, D. and Vassileva, J. (2010). Recommendations in online discussion forums for e-learning systems. IEEE transactions on learning technologies., 3(2): 165-176. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5288522

[7] Borah, N., Baruah, U., Thylore Ramakrishna, M., Kumar, V. V., Dorai, D. R., & Rajkumar Annad, J. (2023). Efficient Assamese Word Recognition for Societal Empowerment: A Comparative Feature-Based Analysis. IEEE Access, 11, 82302–82326. https://doi.org/10.1109/access.2023.3301564

[8] Isinkaye, F. O., Folajimi, Y. O., and Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal, 16(3), 261–273. https://doi.org/10.1016/j.eij.2015.06.005

[9] Ricci, F., Rokach, L., and Shapira, B. (2015). Recommender Systems: Introduction and Challenges. Recommender Systems

Handbook, 1–34. doi:10.1007/978-1-4899-7637-6_1 https://link.springer.com/chapter/10.1007/978-1-4899-7637-6_1

[10] JussiKarlgren. (2019). A digital bookshelf: original work on recommender systems. [online] Available at: https://jussikarlgren.wordpress.com/2017/10/01/a-digital-bookshelf-original-work-on-recommender-systems

[11] Venkatesan, V. K., Ramakrishna, M. T., Batyuk, A., Barna, A., & Havrysh, B. (2023). High-Performance Artificial Intelligence Recommendation of Quality Research Papers Using Effective Collaborative Approach. Systems, 11(2), 81. https://doi.org/10.3390/systems11020081

[12] Kerkiri, T., Manitsaris, A., and Mavridou, A. (2007). Reputation Metadata for Recommending Personalized e-Learning Resources. Second International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007), 110-115 https://ieeexplore.ieee.org/document/4414396

[13] Liu, F., and Shih, B. (2007). Learning Activity-Based E-Learning Material Recommendation System. Ninth IEEE International Symposium on Multimedia Workshops (ISMW 2007), 343-348. https://ieeexplore.ieee.org/document/4475993

[14] Kim, K., and Ahn, H. (2008). A recommender system using GA K-means clustering in an online shopping market. Expert Systems with Applications, 34(2), 1200–1209. https://doi.org/10.1016/j.eswa.2006.12.025

[15] Dakhel, G.M., and Mahdavi, M. (2011). A new collaborative filtering algorithm using K-means clustering and neighbors' voting. 2011 11th International Conference on Hybrid Intelligent Systems (HIS), 179-184. https://ieeexplore.ieee.org/document/6122101

[16] Klašnja-Milićević, A., Ivanović, M., Vesin, B., and Budimac, Z. (2017). Enhancing e-learning systems with personalized recommendation based on collaborative tagging techniques. Applied Intelligence, 48(6), 1519–1535. https://doi.org/10.1007/s10489-017-1051-8

[17] Dwivedi, P., Kant, V., and Bharadwaj, K. K. (2017). Learning path recommendation based on modified variable length genetic algorithm. Education and Information Technologies, 23(2), 819–836. https://doi.org/10.1007/s10639-017-9637-7

[18] Basheer, S., Anbarasi, M., Sakshi, D. G., & Vinoth Kumar, V. (2020). Efficient text summarization method for blind people using text mining techniques. International Journal of Speech Technology, 23(4), 713–725. https://doi.org/10.1007/s10772-020-09712-z

[19] Katarya, R. (2018). Movie recommender system with metaheuristic artificial bee. Neural Computing and Applications, 30(6), 1983–1990. https://doi.org/10.1007/s00521-017-3338-4

[20] Djellali, H., Djebbar, A., Zine, N. G. and Azizi, N. (2018). Hybrid Artificial Bees Colony and Particle Swarm on Feature Selection. Computational Intelligence and Its Applications., 522: 93-105 https://inria.hal.science/hal-01913902

[21] Bagchi, M. (2020). Conceptualizing a Library Chatbot using Open Source Conversational Artificial Intelligence. DESIDOC Journal of Library and Information Technology, 40(06), 329-333. https://doi.org/10.14429/djlit.40.06.15611

[22] Jagadeesan, S., and Subbiah, J. (2020). Real-time personalization and recommendation in Adaptive Learning Management System. J. Ambient Intell. Humaniz. Comput., 11, 4731-4741. https://link.springer.com/article/10.1007/s12652-022-03996-6

[23] Kulkarni, P., Sunil Rai and Rohini, S., (2020). Recommender System in eLearning:A Survey. Proceeding of International Conference on Computational Science and Applications, 13: 119-126, https://doi.org/10.1007/978-981-15-0790-8

[24] Patro, S. G. K., Mishra, B. K., Panda, S. K., Kumar, R., Long, H. V., Taniar, D., and Priyadarshini, I. (2020). A Hybrid Action-Related K-Nearest Neighbour (HAR-KNN) Approach for Recommendation Systems. IEEE Access, 8, 90978–90991. https://doi.org/10.1109/access.2020.2994056

[25] Monsalve-Pulido, J., Aguilar, J., Montoya, E., and Salazar, C. (2020). Autonomous recommender system architecture for virtual learning environments. Applied Computing and Informatics, ahead-of-print(ahead-of-print). https://doi.org/10.1016/j.aci.2020.03.001

[26] Shi, D., Wang, T., Xing, H., and Xu, H. (2020). A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning. Knowledge-Based Systems, 195, 105618. https://doi.org/10.1016/j.knosys.2020.105618

[27] Nguyen, S. T., Kwak, H. Y., Lee, S. Y., and Gim, G. Y. (2021). Featured Hybrid Recommendation System Using Stochastic Gradient Descent. International Journal of Networked and Distributed Computing, 9(1), 25. https://doi.org/10.2991/ijndc.k.201218.004

[28] Yaman, A., Sartono, B., Indrawati, A., Kartika, Y., and Soleh, A. M. (2022). Automated Multi Label Classification on Fertilizer Themed Patent Documents in Indonesia. DESIDOC Journal of Library and Information Technology, 42(4), 218-226. https://doi.org/10.14429/djlit.42.4.17733

[29] Young, S. W. H. (2017). Digital Library User Research Data [dataset]. Scholarworks.montana.edu. https://doi.org/10.15788/M2W88P

[30] Cuneiform Digital Library Initiative. (n.d.). www.kaggle.com. Retrieved July 8, 2022, from https://www.kaggle.com/datasets/mylesoneill/cuneiform-digital-library-initiative

[31] Book Recommendation Dataset. (n.d.). www.kaggle.com. https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset

[32] Sain, S. R., and Vapnik, V. N. (1996). The Nature of Statistical Learning Theory. Technometrics, 38(4), 409. https://doi.org/10.2307/1271324

[33] Jayasuruthi, L., Shalini, A., & Kumar, V. V. (2018). Application of Rough Set Theory in Data Mining Market Analysis Using Rough Sets Data Explorer. Journal of Computational and Theoretical Nanoscience, 15(6), 2126–2130. https://doi.org/10.1166/jctn.2018.7420

[34] Guenounou, O., Dahhou, B., and Chabour, F. (2015). TSK fuzzy model with minimal parameters. Applied Soft Computing, 30, 748–757. https://doi.org/10.1016/j.asoc.2015.02.017

[35] Mahesh, T. R., Vinoth Kumar, V., Vivek, V., KarthickRaghunath, K. M., and Sindhu Madhuri, G. (2022). Early predictive model for breast cancer classification using blended ensemble learning. International Journal of System Assurance

Engineering and Management. https://doi.org/10.1007/s13198-022-01696-0

[36] Raghunath, K.M., and Rengarajan (2013). Investigation of Faults, Errors and Failures in Wireless Sensor Network: A Systematical Survey, International Journal of Advanced Computer Research, 3(3), 151-163 https://accentsjournals.org/paperInfo.php?journalPaperId=595 and countPaper=9219

[37] Ahmed, S. T., Kumar, V. V., & Kim, J. (2023). AITel: eHealth Augmented-Intelligence-Based Telemedicine Resource Recommendation Framework for IoT Devices in Smart Cities. IEEE Internet of Things Journal, 10(21), 18461–18468. https://doi.org/10.1109/jiot.2023.3243784

Edited by: Dhilip Kumar V
Special issue on: Unleashing the Power of Edge AI for Scalable Image and Video Processing
Received: Nov 19, 2023
Accepted: May 3, 2024