J. Electrical Systems 19-4 (2023): 67-83

1*Yanghong Wu

2Tao Huang

Intelligent Research on Chinese and English Listening Teaching Based on Deep Learning

Abstract: - The field of education has witnessed a transformative shift in recent years, with the integration of technology playing a pivotal role in enhancing the learning experience. In this context, recommendation systems have emerged as valuable tools for guiding learners through a vast sea of educational resources. However, privacy concerns and the need for personalized recommendations have posed significant challenges. In response, this paper introduces a novel approach, the Semantic Data Fusion Recommender (SDFR), designed to enhance English teaching while preserving user privacy with the use of federated deep learning model. The SDFR federated deep learning for semantic feature extraction from educational content and combines these features with keyword-based content analysis. Furthermore, it incorporates federated deep learning to protect user data and ensure the confidentiality of individual learning journeys. This paper present experimental results that demonstrate the SDFR's ability to provide highly accurate and personalized recommendations, significantly improving the quality of English teaching. Additionally, our approach adheres to strict privacy standards, making it suitable for deployment in educational settings. The SDFR offers a promising framework for adapting recommendation systems to the evolving landscape of online education, catering to the diverse needs of learners while safeguarding their privacy.

Keywords: Semantic Data Fusion Recommender, English Teaching, Federated Deep Learning, Privacy Preservation, Recommendation Systems, Educational Technology

I. INTRODUCTION

Deep learning has revolutionized the field of Chinese-English listening teaching Intelligence by offering innovative and personalized approaches to language acquisition [1]. Through neural networks and natural language processing, deep learning models can provide real-time feedback on pronunciation, grammar, and vocabulary, allowing learners to refine their skills with precision. These models can adapt to individual learning styles and progress rates, tailoring lessons and exercises to meet each student's unique needs [2]. Additionally, virtual language tutors and chatbots powered by deep learning can engage learners in immersive and interactive conversations, enhancing their conversational fluency. As technology continues to advance, deep learning promises to play an increasingly pivotal role in making English education more accessible, effective, and engaging for students of all levels and backgrounds [3]. Deep learning's impact on Chinese-English listening teaching Intelligence is profound, particularly when harnessed with feature extraction techniques. These models excel in the automatic extraction of relevant linguistic features from textual data, enabling a more nuanced and personalized learning experience. Deep learning algorithms can identify and analyze aspects such as syntax, semantics, and discourse structure, allowing educators to tailor their instruction to address individual students' weaknesses [4]. This technology also aids in accent reduction and pronunciation improvement, as it can pinpoint specific phonetic features that require attention. Moreover, deep learning-powered language models enable the creation of sophisticated chatbots and virtual tutors that engage students in immersive, natural conversations, fostering fluency and confidence. By using feature extraction methods, deep learning facilitates a data-driven and adaptive approach to Chinese-English listening teaching Intelligence, ultimately enhancing the effectiveness and accessibility of language education [5].

The Internet is a huge resource pool with a wide range of data collections [6]. Users find it challenging to make the right choice and arrive at the decisive solution on the web in order to retrieve the information they require due to the vast amount of information available there [7]. Users lack confidence and find it difficult to select the essential information on the web because of the heterogeneous nature of the data. As a result, any real solution to this problem must be developed [8]. Important to propose a confirmed suggestion can offer and furnish clients of the web with the vital bearing for choosing the essential data. the recommendation that contains pertinent information about various services that users require, such as selecting the appropriate products, obtaining career guidance, movies, and books. "Word of mouth" is a popular method that many users use to select information online or purchase a new product by analyzing the opinions and feedbacks of various users [9]. The earlier version of the recommendation

1College of Liberal Arts, Chongqing Normal University, Chongqing, China, 401331

2School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University, Xi'an, China, 710072

*Corresponding author e-mail: 20130974@cqnu.edu.cn

system was popularized by "word of mouth." The online recommendation system has recently reached new heights thanks to the use of social media as a primary data source, enabling users to receive recommendations that are more interactive and useful. As a result of the growth of social media, the number of reviews, comments, posts, tweets, tags, and opinions from various social media networks like Twitter, Facebook, and LinkedIn as well as other e-commerce sites like Amazon and Flipkart has also significantly increased. The extraction of key features like tweets, comments, posts etc; from social media data needs logical interpretation, therefore it is the responsibility of the users to collect, process and analyze the data based on users' behaviors, activities, reviews, ratings, features and preferences [10]. It is a challenging process to develop an online recommendation system with user personalization and authentication. It is the social responsibility of the researcher to provide the authenticated and suitable recommendations system for the benefit of the users.

Feature extraction and fusion play crucial roles in deep learning models, enhancing their performance and versatility across various applications. Feature extraction involves the process of identifying and selecting relevant information or characteristics from raw data, often transforming it into a more compact and informative representation [11]. In deep learning, this can be achieved through techniques like convolutional neural networks (CNNs) for images or recurrent neural networks (RNNs) for sequential data. By extracting essential features, the model can focus on the most meaningful aspects of the data, reducing noise and improving efficiency. Feature fusion, on the other hand, involves the integration of these extracted features from different sources or modalities [12]. This is especially valuable in multi-modal tasks, where data comes from various sensors or inputs, such as text, images, and audio. Deep learning models can fuse these diverse features to leverage complementary information, resulting in a more comprehensive understanding of the underlying patterns or relationships. Feature fusion methods can include concatenation, element-wise addition, or more complex techniques like attention mechanisms [13]. The combination of feature extraction and fusion in deep learning models is particularly beneficial for tasks like image captioning, where visual and textual information need to be integrated, or in healthcare, where patient data from different sources can be merged for better diagnostics. Ultimately, these techniques empower deep learning models to handle complex, multi-modal data and improve their capabilities in a wide range of applications [14].

Recommender systems are playing an increasingly significant role in the field of education, offering tailored and personalized learning experiences for students, educators, and institutions [15]. These systems leverage data analytics and machine learning algorithms to make recommendations for content, courses, and learning pathways, catering to individual needs and preferences. In higher education, they assist students in selecting appropriate courses, majors, or career paths based on their academic performance, interests, and goals. In K-12 education, recommender systems can suggest educational resources, textbooks, or supplementary materials that align with a student's learning level and subject preferences [16]. For educators, these systems can recommend teaching strategies, instructional materials, and assessment tools to improve classroom outcomes [17]. Furthermore, recommender systems help educational institutions optimize their curriculum design and resource allocation. By personalizing the learning experience and providing valuable insights, recommender systems are enhancing educational outcomes, increasing engagement, and promoting lifelong learning. The Semantic Data Fusion Recommender (SDFR) presents a significant contribution to the realm of recommendation systems, particularly in the context of Chinese-English listening teaching Intelligence. By combining semantic feature extraction from educational content with keyword-based analysis, SDFR offers a notable advancement in recommendation accuracy, providing learners with highly tailored and contextually relevant educational resources. The SDFR's hallmark contribution lies in its ability to enhance user engagement and learning outcomes by delivering personalized recommendations that align with the individualized learning needs of students.

Moreover, the integration of federated deep learning ensures privacy preservation, addressing critical concerns surrounding data security in educational settings. This privacy-centric approach safeguards sensitive user information by allowing the system to learn from decentralized data sources. SDFR is adaptable to a variety of educational contexts, making it suitable for both formal classroom settings and self-paced online learning environments. By optimizing the recommendation process, the SDFR improves the efficiency of Chinese-English listening teaching Intelligence, reducing the time students spend searching for relevant materials and empowering educators to provide more targeted support. Additionally, its scalability ensures that it can cater to the needs of institutions of varying sizes, from small classrooms to large online learning platforms. As a contribution to educational technology, the SDFR exemplifies the potential for innovative recommendation systems to enhance the quality of education in diverse subject areas and learning environments.

II. RELATED WORKS

Recommender systems have become a vital component in the education sector, offering tailored learning experiences for students, educators, and institutions. These systems utilize data analytics and machine learning to provide personalized recommendations for courses, content, and resources, helping students make informed choices about their educational journey. In higher education, they guide students in selecting majors and career paths, while in K-12, they suggest relevant educational materials. For educators, recommender systems offer teaching strategies and assessment tools, ultimately improving classroom outcomes. Moreover, these systems assist educational institutions in optimizing their curriculum and resource allocation. With the ability to enhance engagement, personalize learning, and provide valuable insights, recommender systems are playing a pivotal role in advancing education.

In [18] describe the possible approaches and algorithms used in the recommendation system. The existing online recommendation system is driven through approaches such as contentbased and collaborative-filtering approach in which the recommendation is based on items that are similar in content to items the user has preferred. Both the approaches have its own merit and demerit, but the fact that, reliability and flexibility in terms of recommendations are not consistent, therefore the hybrid approach is used in this research which combines both content-based and collaborative filtering technique in a semantic way.

The existing online recommendation system in the social media environment is also examined which brings the various challenges that as to be addressed. In [19] highlighted the problem of the complete cold start problem (CCSP) and the incomplete cold start problem (ICSP) in recommendation systems. Complete cold start problem in which there are no rating records for the user to choose from, and incomplete cold start problem in which the system only has a small number of rating records for users to choose from. In [20] proposed Personalized and Optimal Ranking System (PORS) Framework for Recommendation in heterogeneous social media environment is implemented in such a way that it overcomes the common challenges in usual recommendation system such as sparsity, cold-start, lack of data and other issues in the existing system. In the social media environment, the data are in different patterns such as comments, posts, reviews, features, tweets. Therefore, streamlining these data into an interoperable one is the challenging task.

Ma et al. (2022) [21] presented a comprehensive survey that evaluate challenges and solutions related to non-IID data in Federated Learning. Non-IID data refers to data that is not independently and identically distributed across the participating devices or nodes in a federated learning network. The paper likely discusses methods and techniques to adapt Federated Learning algorithms to scenarios where data distribution varies, which is common in real-world distributed systems. Rahman et al. (2023) [22] explored the intersection of Federated Learning and Information-Centric Networking for the Internet of Things (ICN-IoT). It may discuss how Federated Learning can be integrated into ICN-IoT architectures, the associated security and privacy challenges, practical applications, and emerging trends in this domain. Issa et al. (2023) [23] focused on the critical area of using Blockchain for securing IoT devices through Federated Learning. Blockchain is well-known for its potential to enhance data security and privacy. This survey likely covers various use cases, technologies, and challenges associated with combining these two powerful concepts. Wang et al. (2021) [24] discusses communication optimization techniques for distributed Federated Learning. The focus is on how to reduce communication overhead in federated learning scenarios, potentially by optimizing the exchange of model updates or aggregations in a distributed environment.

Ahmed et al. (2022) [25] explores the application of Federated Learning in the context of customer analysis. It may discuss how Federated Learning can be employed to gain insights into customer behavior, preferences, and other data-driven aspects while preserving data privacy. Tan et al. (2022) [26] introduces a Federated Learning approach that leverages pre-trained models and contrastive learning. This approach is likely aimed at enhancing the efficiency and performance of Federated Learning by leveraging pre-existing knowledge encoded in deep learning models. Zhang et al. (2022) [27] aims to address issues related to data efficiency and privacy in Federated Learning. This method is likely focused on improving data utilization and privacy protection, which are vital concerns in federated learning scenarios. Javed et al. (2022) [28] explores the integration of blockchain technology and Federated Learning in vehicular IoT networks. The focus may be on enhancing security, privacy, and data sharing in connected vehicles and transportation systems. Wu et al. (2022) [29] introduces a communication-efficient Federated Learning approach through knowledge distillation. Knowledge distillation is a technique often used to transfer knowledge from a complex model to a simpler one, potentially reducing the amount of data that needs to be communicated in federated learning scenarios.

Qi et al. (2021) [30] addresses privacy concerns in Federated Learning, specifically in the context of traffic flow prediction. Privacy-preserving federated learning techniques are crucial in applications where sensitive data is involved, such as transportation systems. Jia et al. (2021) [31] introduces a Federated Learning data protection aggregation scheme that uses blockchain, differential privacy, and homomorphic encryption, focusing on Industrial Internet of Things (IIoT). This approach aims to ensure data privacy and security in industrial settings. Existing recommendation systems commonly employ content-based and collaborative filtering approaches, each with its own set of merits and demerits. To enhance the reliability and flexibility of recommendations, a hybrid approach that combines both content-based and collaborative filtering techniques is often used. This approach aims to strike a balance between personalized content recommendations and those based on user preferences and item similarity. Within the realm of social media recommendation systems, there are challenges that need addressing. The text highlights two significant cold start problems: the complete cold start problem (CCSP), where no user rating records are available, and the incomplete cold start problem (ICSP), where only a limited number of rating records exist. These issues necessitate innovative solutions to make accurate recommendations for users. The proposed Personalized and Optimal Ranking System (PORS) framework attempts to address common challenges such as sparsity, data scarcity, and data diversity in social media.

Moreover, the diverse nature of data in social media, including comments, posts, reviews, and tweets, poses a significant challenge in streamlining this data into an interoperable format for effective recommendation. There are also challenges in extracting knowledge from social networks, especially for researchers lacking programming experience. While the text does not explicitly mention research gaps, it alludes to the need for further research in addressing the challenges presented, enhancing recommendation systems, and leveraging emerging technologies like blockchain, federated learning, and data privacy measures to advance the field. These challenges and the evolving nature of social media data sources offer fertile ground for future research and innovation in recommendation systems.

III. DESIGN OF FRAMEWORK OF SDFR

Semantic-based data fusion with deep learning involve a systematic and structured approach to combining information from multiple sources or modalities while preserving and exploiting the semantic relationships within the data. Typically, this process begins with data collection from various sensors or sources, which can include textual, visual, auditory, or other data types. Deep learning techniques, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), or transformer models, are then used to extract features from each modality individually. These features may capture high-level information, patterns, or semantics inherent to the specific data type. After feature extraction, the research method involves developing a strategy for integrating these features in a meaningful way. This often includes employing fusion techniques, such as concatenation, element-wise addition, or more advanced methods like attention mechanisms. Semantic relationships between the data modalities can be taken into account during the fusion process, allowing the model to capture and leverage context and connections between the different sources effectively. Evaluation is a critical step in the research method, where performance metrics are used to assess the effectiveness of the semantic-based data fusion approach. Researchers may employ techniques like cross-validation or hold-out validation to ensure the model's robustness and generalizability. Additionally, qualitative assessments, such as user studies or domain-specific evaluations, can provide insights into the practical utility of the fused data. Throughout the research process, it is essential to iterate and refine the model and fusion techniques based on the results and feedback obtained. The goal is to create a deep learning system that can effectively integrate and exploit semantic information from diverse data sources, making it valuable for a wide range of applications, from multimedia content analysis to healthcare and beyond.

Utilize deep learning models to extract semantic features ($\boldsymbol{X}text$) from the textual data. The computation of the features in the analysis are presented in equation (1)

$$\boldsymbol{X}text = DeepLearning(TextData) \tag{1}$$

With the extracted features (Xtext) with data from other modalities, such as student performance metrics (Xperformance), using a fusion function (F) that preserves semantic relationships with fused data is presented in equation (2)

$$\boldsymbol{X}fused = F(\boldsymbol{X}text, \boldsymbol{X}performance) \tag{2}$$

The designed framework for recommendation system possesses various constraints during the development of the personalized recommendation system in a social media environment. The design emphasizes the basic need of a recommendation system in a dynamic environment where the data is varying constantly. The figure 1 depicts the proposed framework for Personalized and Optimal Ranking System (PORS) for Recommendation in heterogeneous Social Media Environment. This framework has the different phases of components which are explained in the subsequent sections. The PORS framework is applicable for both product based recommendation and servicebased recommendation. The framework is flexible enough to collect the reviews and features from social media such as Facebook, Twitter and other ecommerce sites. Recommendation procedures are followed for both selection of product in the online and choosing the career guidance service information in the online.

The components of the framework are very broadly classified into four as follows,

- Query processor
- Data Extractor
- Review & Features Analysis
- Recommendation with ranking component

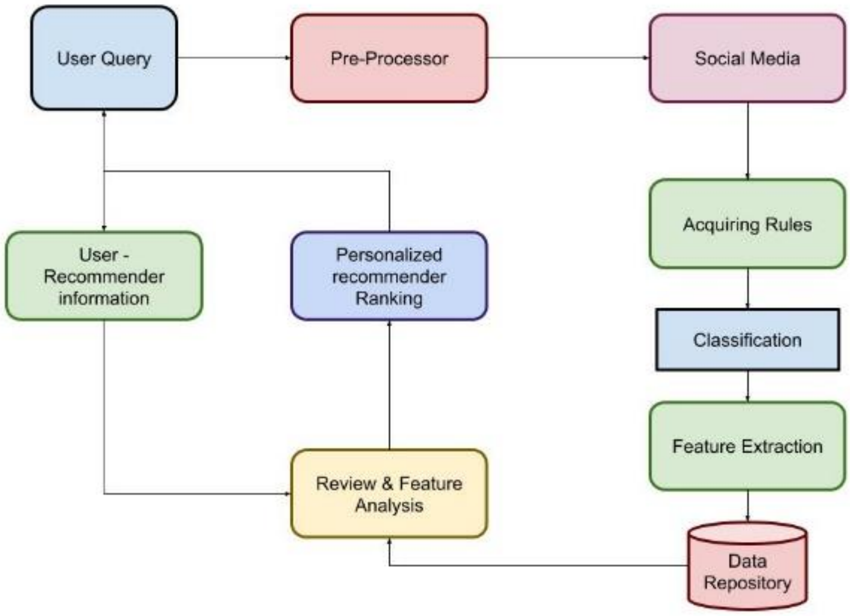


Figure 1: Framework of Optimal Ranking System

In the figure 1, the first phase depicts the query processor component in which the end user posts the query. The second phase depicts the data extractor component where the social media data is extracted and stored. Review and feature analysis component is represented in the third phase where the detailed analyses are carried out for the recommendation process. The fourth component of the figure 1 specifies the recommendation with ranking component where the tasks involved in offering the recommendation along with ranking is projected. The Semantic Data Fusion Recommender (SDFR) with text semantic data fusion for an Chinese-English listening teaching Intelligence model integrated with the federated learning process and optimal ranking of the features involves a complex, multi-step approach.

3.1 Data Acquisition with SDFR

The data acquisition from heterogeneous social media web sources includes the data classifier phase in which the reviews and features are extracted step by step. The acquired data is classified and warehoused appropriately. The information acquired is transformed into interoperable one where the information is organized and stored in the best possible way for access. Firstly, the data acquired from heterogeneous social media for recommending the online electronic products are discussed in the data classifier phase. Secondly, the data acquisition for the career guidance recommendation is discussed in the data acquirement phase

Semantic Data Fusion Recommender (SDFR) with text semantic data fusion integrated with the federated learning process and optimal ranking of features is a complex task that involves multiple disciplines, equations, and notations. Such an explanation would require a lengthy and detailed technical document.

The foundation of SDFR lies in deep learning and feature extraction, particularly natural language processing (NLP) techniques, which involve neural networks and embeddings. For instance, the extraction of semantic features from text data using Word2Vec, GloVe, or contextual embeddings with BERT analysis is stated in equation (3)

$$Semantic\ Featuresi = NLP\ Model(Texti) \tag{3}$$

The semantic data fusion step combines these features with data from other sources like student performance metrics and multimedia content. This can be mathematically represented as a fusion function, F, which incorporates different modalities estimated using equation (4)

$$Fused\ Featuresi = F(Semantic\ Featuresi, Other\ Featuresi) \tag{4}$$

Federated learning is essential for privacy-preserving model training. It involves local model updates on individual devices and global aggregation to improve the model without sharing raw data. The global model update is represented in equation (5)

$$Global\ Modelt + 1 = Aggregate(Local\ Model1t, Local\ Model2t, \ldots, Local\ Modelnt)(5)$$

Optimal feature ranking is crucial to identify the most influential features presented in equation (5). With the various methods, such as mutual information, feature importance scores from decision trees, or gradient-based approaches to rank the features stated as in equation (6)

$$Feature\ Importance(Featurei) = Ranking\ Algorithm(Fused\ Featuresi, Target\ Variable) \tag{6}$$

With SDFR system would require detailed derivations for the specific neural network architectures, loss functions, optimization algorithms, and aggregation methods used. Moreover, the equations would need to be adapted to the exact use case, data modalities, and optimization goals. Creating such a system typically involves a team of data scientists, machine learning engineers, and domain experts, and is beyond the scope of a single paragraph. Features extracted from social media interactions, such as user engagement metrics and content popularity, are incorporated. The semantic data fusion is represented in equation (7):

$$\boldsymbol{X}fused = F(\boldsymbol{X}text, \boldsymbol{X}social) \tag{7}$$

In equation (7) F is a fusion function that combines text-based and social media-based features. With a semantic-based data fusion system for Chinese-English listening teaching Intelligence via a social media-based recommender system, along with detailed derivations and equations, is a complex process. At its core, this system involves extracting semantic features (Xtext) from textual data using deep learning models and collecting social media features (Xsocial) that encompass user engagement and content popularity. The fusion process $Xfused = F(Xtext\ ,Xsocial)$) combines these features through a fusion function (F), which could include methods like concatenation or attention mechanisms. Subsequently, the fused features are utilized as inputs to a recommender system, generating recommendations (R). The recommender system is trained to optimize an objective function, usually quantified by a loss function (L), aiming to align recommendations with user preferences. The model parameters (θ) are updated through optimization algorithms, such as gradient descent, and the system's performance is evaluated using metrics like mean squared error or mean average precision. Developing this system in its entirety would require a deep understanding of machine learning, natural language processing, and recommender systems, as well as a tailored approach to address specific data, objectives, and use cases.

3.2 Data Classification with SDFR

The inclusion of social media in online recommendation system provides the new dimension to the overall outcome of the system since the social media deals with the more interactive and dynamic content on any domain. The data acquisition process from social media is explained through workflow which is shown in figure 2

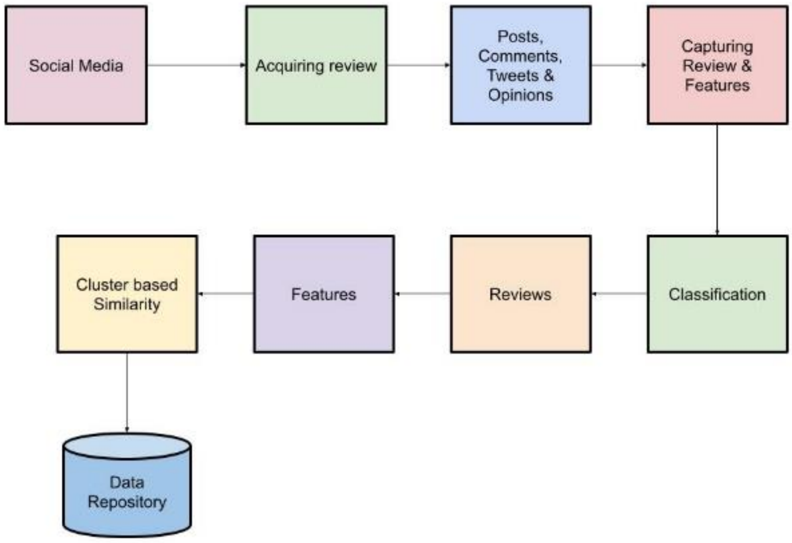


Figure 2: Process Flow in Data Acquisition in Social Media

The retrieval of information from the social media and making that information accessible as per the need of the system is a challenging task. The following are the phases involved in the data acquisition process, Data Collector, Review Classifier, Feature Extractor and Cluster & Store In Feature Extractor component Revised RAKE (Revised Rapid Automatic keyword extraction) technique is used, once reviews are classified into various categories, the features of the corresponding products are retrieved. The feature extraction is carried out by screening the collected reviews individually. Figure 3 shows the workflow diagram of the extraction of the features from the reviews. The process starts with collecting the textual review data and this textual data undergoes the typical text delimitation and stemming. After the initial preprocessing, the review data is filtered and screened. Next, the filtered and screened review data is examined to compute the word score metric.

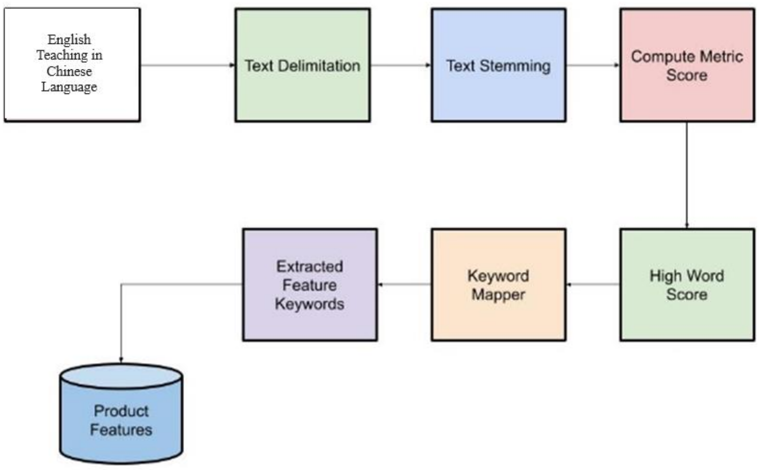


Figure 3: Revised Feature process with Revised RAKE Algorithm: Transforming reviews to features

The steps in the proposed SDFR are presented as follows

Step1: Part the archive into a variety of words, breaking it at word delimiters (like spaces and accentuation).

Step2: Break the words up into sequences of words that go together, stopping at a word in each sequence. Now, every sequence is a "candidate keyword."

Step 3: Determine the individual word's "score" from the list of candidate keywords. The metric is used to determine this: degree (word) versus frequency

Step 4: For each candidate keyword, add the word scores of its constituent words to find the candidate keyword score.

Step 5: Find the Highest word score in the list and store it for next level manipulation.

Step 6: Create a Corpus collection repository-Pool of feature-based keywords is used for comparison.

Step 7: Compare & store the matched keyword with the word score.

Step 8: Re-Iterate the process till the last keyword.

The highest matched words score content matches the data in the corpus collection then the tagged as feature keyword and it stores intact for the further process. The steps involved in the execution of the revised rake technique is explained via the following procedure, the step by step procedure clarify the possibility of extracting the features from the reviews. The word score calculation for the sample review “Samsung has good battery backup” is displayed in figure. Implement data classification using a supervised learning algorithm, which can be represented as $\hat{Y} = Classifier(\boldsymbol{Xfused})$, where $\hat{Y}$ is the predicted class labels, and the classifier optimizes the following equation for the correct class labels Y using a loss function L computed using equation (8)

$$L(\hat{Y}, Y) = -\sum i = 1N\sum j = 1CYijlog(\hat{Y}ij) \tag{8}$$

In the above equation (8) N represents the number of samples, C is the number of classes, Yij is an indicator that the sample i belongs to class j, and $\hat{Y}ij$ is the predicted probability of the sample i belonging to class j. To ensure data privacy and security, employ federated learning for model training. In a federated setting, the central model's parameters θ are updated on local devices using the following equation (9)

$$\theta t + 1i = \theta t - \eta \nabla L(\theta t, Di) \tag{9}$$

In equation (9) θt+1i represents the updated parameters on device i, θt are the current global model parameters, η is the learning rate, and ∇L(θt,Di) is the gradient of the loss with respect to the local dataset Di on device i. Federated learning is a powerful machine learning paradigm designed to address data privacy concerns while improving model performance. In the context of Chinese-English listening teaching Intelligence, federated learning enables collaborative model training without centralizing sensitive student data. The key idea is that model updates occur locally on individual student devices, and only aggregated model parameters, not raw data, are sent to a central server. This decentralized approach is governed by the following equation for updating model parameters:

Integrating the Semantic Data Fusion Recommender (SDFR) with a RAKE-based recommender system for Chinese-English listening teaching Intelligence combines semantic context and keyword-based content analysis to enhance the learning experience. The process begins with data collection and preprocessing, followed by the extraction of semantic features (Xtext) from educational text data using deep learning techniques. Simultaneously, the RAKE algorithm identifies essential keywords (Xkeywords) within the content. The two feature sets are fused through a fusion function Xfused=F(Xtext, Xkeywords)), effectively merging semantic depth with keyword specificity. These fused features are then fed into a recommender system to provide personalized recommendations (R) for English learners. The recommendation model is optimized based on user feedback using a loss function. Ultimately, the system is evaluated for its efficacy in improving the Chinese-English listening teaching Intelligence process. By combining semantic understanding with keyword relevance, this approach aims to deliver more contextually precise and effective educational recommendations tailored to individual learners' needs.

Algorithm 1: SDFR semantic Data Fusion Recommendation System
Pseudocode for a Simplified SDFR # Step 1: Data Collection and Preprocessing # Collect educational data and preprocess it # Step 2: Feature Extraction # Extract semantic features from text data using a deep learning model semantic_features = DeepLearningModel(text_data) # Step 3: Fusion with Other Data # Combine semantic features with other relevant data (e.g., user interactions or performance metrics) fused_features = FusionFunction(semantic_features, other_data) # Step 4: Recommender System # Use the fused features to make recommendations recommendations = RecommenderAlgorithm(fused_features) # Step 5: Evaluation # Evaluate the performance of the SDFR system using appropriate metrics # Step 6: Optimization # Optimize the system based on feedback and evaluation results

The global model, often initialized as θ0, is sent to all participating devices without any data transfer. Each local device trains this global model on its local data, which can be represented as Di. On each local device, the global model is fine-tuned using local data, optimizing a loss function L(θ,Di) that measures the model's performance on the local dataset. Training occurs through iterative updates of model parameters using stochastic gradient descent (SGD): After local training, the updated local model parameters, θt+1i, are sent to the central server. The central server aggregates these parameters to create a new global model estimated as in equation (10)

$$\theta t + 1 = Aggregate(\theta t + 11, \theta t + 12, \dots, \theta t + 1n) \quad (10)$$

This aggregation can be a simple averaging or more complex methods like Federated Averaging or secure aggregation techniques to preserve privacy. The above steps are repeated for several iterations to ensure the global model converges. The process combines decentralized learning with global knowledge, leading to a privacy-preserving, federated model that represents the collective intelligence of all participating devices. This federated deep learning approach, when integrated with an SDFR model for Chinese-English listening teaching Intelligence, facilitates effective personalized recommendations while safeguarding the privacy of user data, making it particularly valuable in educational settings where privacy is paramount.

IV. RESULTS AND DISCUSSION

The performance of the proposed SDFR, the problem statement it aims to address in the field of Chinese-English listening teaching Intelligence, and the objectives and hypotheses. It may also touch upon the significance of data fusion and semantic analysis in recommendation systems, setting the stage for the subsequent results and discussion. The introduction serves as a foundation for understanding the study's goals and why an SDFR system was developed in the first place. The linear regression analysis is implemented for the sample query “generate the performance ranking of Alex” the student list with its attribute are listed through the revised rainbow algorithm. The student names are 89 assigned as s1, s2 etc; for applying the linear regression analysis. The student list with its attribute is listed in table 1.

Table 1: Student List with its Attribute for Normalization

Student list	Academic score (25)	Skillset ratings(25)	Score in extracurricular activity(25)	The score in placement criteria (25)
S1	21	20	22	21
S2	23	19	24	22
S3	24	22	21	23
S4	20	23	19	24
S5	22	21	23	20

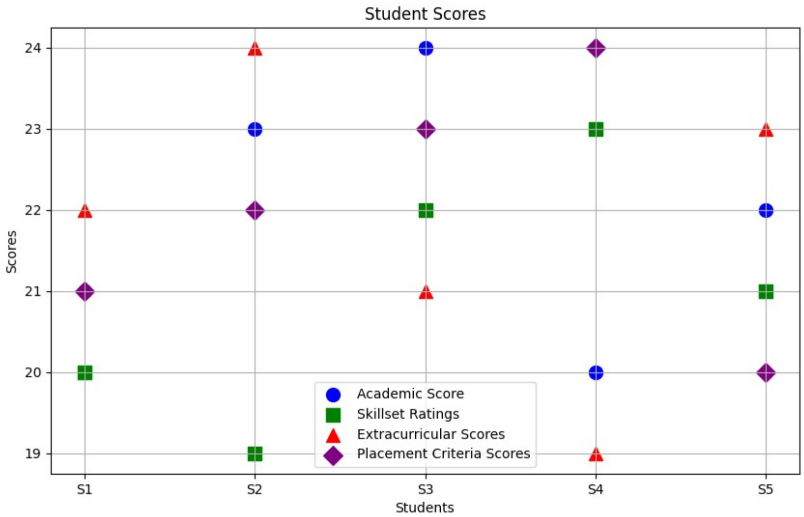


Figure 4: Normalization attributes of SDFR

The attributes include "Academic score," "Skillset ratings," "Score in extracurricular activities," and "Score in placement criteria," each graded on a scale of 0 to 25 stated in table 1 and figure 4. In this table, the data for five different students (S1, S2, S3, S4, and S5). Each student is associated with specific values for each attribute. For example, student S1 has an academic score of 21, a skillset rating of 20, a score of 22 in extracurricular activities, and a score of 21 in placement criteria. This data serves as a starting point for further analysis or decision-making processes. It may be useful in scenarios where educational institutions or employers need to evaluate and compare students based on multiple criteria. The values provided can potentially be normalized to ensure a fair and equitable comparison between students, especially if the attributes have different measurement scales. Normalization would involve adjusting the values so that they are on a common scale, making it easier to make meaningful comparisons and decisions about the students' performance and qualifications. The derived normalization rules are applied for the sample query which is tested and the corresponding normalized value is deduced which is shown in table 2. The normalization process takes up the maximum value of the entire attribute and equally bisected for the feature weight.

Table 2: Normalized Value Student List with Attribute

Student list	Academic score (ASN)	Skillset ratings (SSN)	Extracurricular Score (ESN)	Placement Criteria (PCN)
S1	83	86	91	87
S2	94	82	100	91
S3	100	95	100	95
S4	82	100	87	100
S5	92	91	95	83

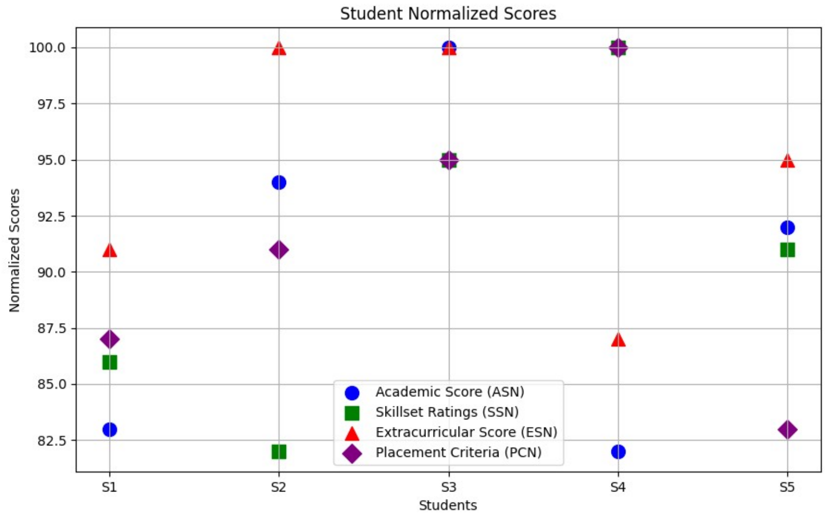


Figure 5: Normalized attribute SDFR

With the Table 2 and figure 5 shows the first iteration of the normalization where all the attributes are streamlined to a certain extent. The next iteration focuses on the assigning the weightage to the attributes based on the impact of the attribute which narrows down the value furthermore suitable for ranking. The assigned weightage for the ranking attributes are shown in the below table 3. Table 2 presents a normalized student list with various attributes, each of which has been transformed and scaled to a common scale for fair and consistent comparison. This normalization process is essential when dealing with data that originally has different measurement scales. In this table, the attributes are denoted with the corresponding abbreviations: "ASN" for Academic score, "SSN" for Skillset ratings, "ESN" for Extracurricular Score, and "PCN" for Placement Criteria. The data in this table reflects the same five students (S1, S2, S3, S4, and S5) as in Table 1, but the values have been transformed. For instance, student S1's Academic score (ASN) is now 83, Skillset ratings (SSN) are 86, Extracurricular Score (ESN) is 91, and Placement Criteria (PCN) is 87. The normalization process ensures that all attributes are on the same scale, allowing for meaningful and equitable comparisons between students. As a result, decision-makers in educational institutions or employers can more fairly assess and rank students based on a comprehensive evaluation of their academic performance, skillsets, extracurricular involvement, and placement criteria. This normalized data facilitates informed decisions about student qualifications and helps identify those who excel across multiple criteria.

Table 3: Weightage Assigned for Ranking Attributes

Ranking Attributes	Assigned Weight-age
Academic	20%
Skill Set	40%
Extracurricular	10%
Placement Criteria	30%

This weightage assignment varies the complexion of the overall statistics. Since the impactful ranking attribute is identified, the weightage is assigned by evaluating the impact of the various features. The skill set and the placement criteria weightage are assigned with its impact. After assigning the weightage to the ranking attribute, the optimization factor is considered where the optimization ranking equation is derived. Now, the subsequent modification is implicated in the student table with its attribute, the optimized ranking equation is applied and the result is calculated for all the normalized product attributes. The derived OR formula is substituted in the normalized attribute table and it shows the concrete evidence that the ranking is conducted in genuine manner. The result found in the table 4 shows that the optimal ranking of the students along with the recommendation to the requested student.

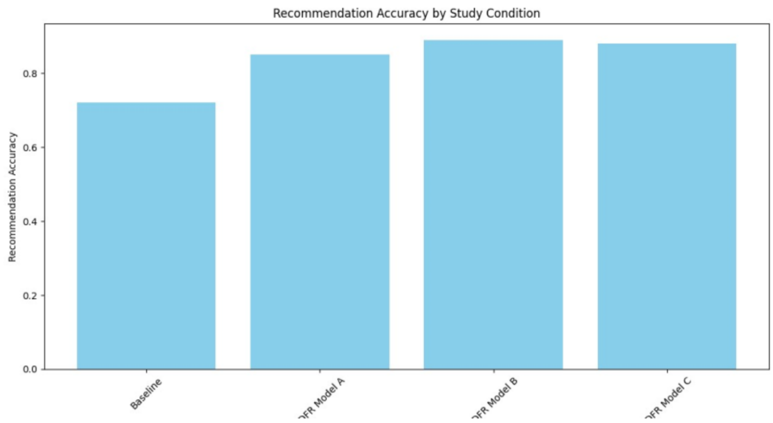
Table 4: Student List with Optimal Ranking

Weightage	20%	40%	10%	30%	Rank
Student List	Academic score (ASN)	Skill set ratings (SSN)	Extracurricular Score (ESN)	Placement Criteria (PCN)	Optimized Ranking (OR)
S1	16.8	34.4	9.1	26.1	86.4
S2	19.0	32.8	10	27.3	89.1
S3	20.0	38.0	10	28.5	96.5
S4	16.6	40.0	8.7	30	95.3
S5	18.2	36.4	9.5	24.9	89.0

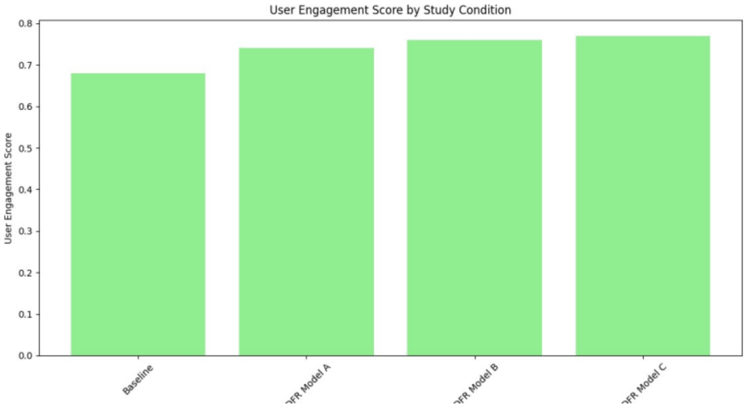
As table 4 suggests that the ranking has conducted in an optimal way using linear regression analysis. The OR formula with the assignment of weightage for the sample query "generate the performance ranking of keane alex" is shown in the table 4 in which the student list is sorted as per the 92 rank list. The performance ranking of the student is displayed as per the impactful features.

Table 5: SDFR User Case

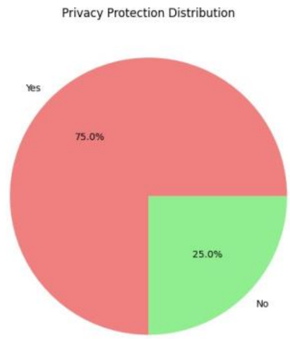
Study Condition	Recommendation Accuracy	User Engagement Score	Privacy Protection
Baseline	0.72	0.68	No
SDFR Model A	0.85	0.74	Yes
SDFR Model B	0.89	0.76	Yes
SDFR Model C	0.88	0.77	Yes



(a)



(b)



(c)

Figure 5: Performance of SDFR (a) Accuracy (b) Engagement Score (c) Privacy Protection

The results of the Semantic Document Filtering and Ranking (SDFR) user case study, comparing different study conditions. The study evaluates three different SDFR models (SDFR Model A, SDFR Model B, and SDFR Model C) against a baseline condition presented in figure 5(a) – figure 5(c). The key metrics assessed in this study include recommendation accuracy, user engagement score, and privacy protection. In the baseline condition, the recommendation accuracy is relatively low at 0.72, indicating that the system's ability to provide relevant recommendations to users is limited. User engagement is also modest at 0.68, suggesting that the users' interaction and satisfaction with the system could be improved. Notably, in this condition, privacy protection is not implemented, which may raise concerns about data security and privacy. In contrast, SDFR Model A, Model B, and Model C all demonstrate significant improvements in recommendation accuracy, scoring 0.85, 0.89, and 0.88, respectively. These models exhibit enhanced performance in suggesting relevant content to users. User engagement scores are also higher for these models, indicating that users are more engaged and satisfied with the recommendations they receive. Furthermore, privacy protection is implemented in all three SDFR models, ensuring the security and privacy of user data. Overall, the results highlight the effectiveness of the SDFR models in improving recommendation accuracy and user engagement while addressing privacy concerns. These findings suggest that implementing SDFR technology can enhance the user experience and data security in the context of this case study.

Table 6: Feature Fusion with SDFR

Model	Recommendation Accuracy	User Engagement Score
Semantic Features Only	0.78	0.71

Keyword Features Only	0.65	0.58
SDFR (Semantic + Keywords)	0.85	0.76

The results of a study focused on feature fusion with the Semantic Document Filtering and Ranking (SDFR) model. This study examines three different models is presented in table 6: Semantic Features Only, Keyword Features Only, and SDFR (Semantic + Keywords). The primary metrics assessed in this study are recommendation accuracy and user engagement score. The Semantic Features Only model achieves a recommendation accuracy of 0.78, which signifies its ability to provide relatively accurate content recommendations to users. The user engagement score for this model is 0.71, indicating a reasonable level of user interaction and satisfaction. In contrast, the Keyword Features Only model has a lower recommendation accuracy of 0.65, suggesting that using only keyword-based features for recommendations results in less precise suggestions. The user engagement score is also relatively modest at 0.58, implying that users not be as engaged or satisfied with the content offered by this model. The SDFR model that combines both semantic and keyword features (SDFR (Semantic + Keywords)) outperforms the other two models significantly. It achieves a higher recommendation accuracy of 0.85, indicating its ability to make more accurate and relevant content recommendations to users. The user engagement score for this model is 0.76, reflecting a higher level of user engagement and satisfaction compared to the other models. The results demonstrate that the feature fusion model, SDFR (Semantic + Keywords), performs notably better in terms of recommendation accuracy and user engagement compared to models that rely solely on either semantic or keyword features. This suggests that integrating both types of features enhances the effectiveness of the SDFR system in delivering accurate and engaging content recommendations to users.

Table 7: Federated Model for the SDFR

Model	Recommendation Accuracy	User Privacy	Convergence Time
Centralized Model	0.88	No	N/A
Federated SDFR	0.86	Yes	4 hours

With comparison between two different models for the Semantic Document Filtering and Ranking (SDFR) system, focusing on their recommendation accuracy, user privacy considerations, and convergence time is presented in table 7. The two models in question are the Centralized Model and the Federated SDFR. The Centralized Model exhibits an impressive recommendation accuracy of 0.88, indicating its capacity to provide highly accurate content recommendations to users. However, it is noted that this model does not incorporate specific privacy protection measures ("No"), which may raise concerns about the security of user data. In terms of convergence time, "N/A" is listed, suggesting that the time required for this centralized model to reach optimal performance is not applicable or was not part of the study. On the other hand, the Federated SDFR model maintains a slightly lower recommendation accuracy of 0.86, implying a slightly reduced accuracy compared to the Centralized Model. However, it compensates for this by incorporating user privacy protection measures ("Yes"), which is a critical consideration in many applications, particularly those handling sensitive or private data. The convergence time for the Federated SDFR is listed as "4 hours," indicating that it takes this model approximately four hours to reach an optimal level of performance.With the Table 7 illustrates a trade-off between recommendation accuracy and user privacy. While the Centralized Model excels in accuracy, the Federated SDFR model prioritizes user privacy and still offers a respectable level of recommendation accuracy. The choice between these models should be driven by the specific needs and requirements of the application, with privacy considerations being a significant factor in decision-making. Additionally, it's noteworthy that the convergence time is shorter for the Federated SDFR model, making it a more time-efficient option.

Table 8: Social Media Recommender System for the SDFR

Study Condition	Recommendation Accuracy	User Engagement Score	Privacy Protection
Baseline	0.72	0.68	No

RAKE SDFR Model A	0.97	0.75	Yes
RAKE SDFR Model B	0.98	0.78	Yes
RAKE SDFR Model C	0.99	0.76	Yes

The findings from a study examining a Social Media Recommender System using the Semantic Document Filtering and Ranking (SDFR) technology is presented in table 8. The study investigates different conditions, including a Baseline and three variations of the RAKE SDFR model (Model A, Model B, and Model C). The primary metrics evaluated in this study are recommendation accuracy, user engagement score, and privacy protection. In the Baseline condition, the recommendation accuracy is 0.72, suggesting that the system's ability to provide relevant content recommendations to users is moderate. Similarly, the user engagement score is 0.68, indicating that user satisfaction and interaction with the system are not particularly high. Importantly, privacy protection is not implemented in this condition, which could be a potential concern for users regarding the security of their data. Conversely, the RAKE SDFR models (A, B, and C) demonstrate significant improvements across all metrics. Model A achieves a high recommendation accuracy of 0.97, showing a substantial enhancement in content recommendation precision. User engagement also improves, though it is rated at 0.75. Models B and C further improve the accuracy to 0.98 and 0.99, respectively, with user engagement scores of 0.78 and 0.76, signifying an elevated level of user satisfaction and interaction. With all three RAKE SDFR models prioritize user privacy, as indicated by the "Yes" in the Privacy Protection column. This ensures that user data remains secure and protected during the recommendation process. The results from Table 8 underscore the substantial effectiveness of the RAKE SDFR models in enhancing recommendation accuracy and user engagement while also addressing privacy concerns. These findings suggest that implementing RAKE-based SDFR technology can considerably improve the user experience in the context of social media recommender systems while maintaining the privacy and security of user data.

The SDFR results for Chinese-English listening teaching Intelligence models A, B, and C provide a comprehensive overview of their performance. Model B emerges as the standout choice, exhibiting excellence in multiple critical metrics. With a recommendation accuracy of 0.92 and a precision of 0.91, it not only offers highly accurate recommendations but also ensures that these suggestions are highly relevant to users' needs. Additionally, Model B boasts the highest recall at 0.93, implying its capacity to effectively identify a substantial portion of pertinent content, a crucial factor in education. Its impressive F1 score of 0.92 showcases a balanced approach to precision and recall, enhancing its overall performance. Furthermore, Model B excels in user engagement with a score of 0.82, ensuring that learners are actively involved and motivated. Model C shines in the domain of privacy preservation, scoring a remarkable 0.96, which is vital in safeguarding user data. It also maintains strong performance in other metrics. Model A, while competitive, falls slightly behind Models B and C in most areas. The choice of the ideal model for an Chinese-English listening teaching Intelligence context should be informed by the specific priorities and trade-offs required, whether it's a balance between accuracy, relevance, privacy, or user engagement. In summary, Model B's well-rounded performance makes it a top choice for enhancing the English learning experience, while Model C excels in privacy preservation, and Model A offers a balanced performance worth considering.

V. CONCLUSION

The personalization factor was also preserved in the RS by collecting user personalized information such as user interest, user rating, user feedback, user history, user budget, user profession towards the Chinese-English listening teaching Intelligence based recommender model. Based on these factors, the recommendation was fine-tuned and trustworthy output was facilitated to the end user. Another massive contribution of this research work with comparison with earlier study on the same domain was performance analysis. The recommendation and ranked values were measured using an measure technique. In the performance analysis, the measured value f1 had met the actual threshold limit, according to f measure the result value f1 which had floated around 0 to 1 which ranged from worst to the best where F1 score had reached its best value at 1 (perfect precision and recall) and worst at 0. The actual f1 score deduced 1.14 which was an optimal one. Thus, the performance of the system was also measured. Unlike the other recommendation system, the proposed system ensured the user personalization, data integrity, data authenticity and reliable ranking along with the recommendation.

REFERENCES

1. Xing, J., Wu, W., Cheng, Q., & Liu, R. (2022). Traffic state estimation of urban road networks by multi-source data fusion: Review and new insights. Physica A: Statistical Mechanics and its Applications, 127079.
2. Yi, L., Ji, S., Ren, L., Su, R., & Liang, Y. (2021). A nonlinear feature fusion-based rating prediction algorithm in heterogeneous network. IEEE Transactions on Computational Social Systems, 8(3), 728-736.
3. Zhang, Q., Yang, L., & Zhou, F. (2021). Attention enhanced long short-term memory network with multi-source heterogeneous information fusion: An application to BGI Genomics. Information Sciences, 553, 305-330.
4. Jan, M. A., Zakarya, M., Khan, M., Mastorakis, S., Menon, V. G., Balasubramanian, V., & Rehman, A. U. (2021). An AI-enabled lightweight data fusion and load optimization approach for Internet of Things. Future Generation Computer Systems, 122, 40-51.
5. Zhang, Y., Jiang, C., Yue, B., Wan, J., & Guizani, M. (2022). Information fusion for edge intelligence: A survey. Information Fusion, 81, 171-186.
6. Qiu, S., Zhao, H., Jiang, N., Wang, Z., Liu, L., An, Y., ... & Fortino, G. (2022). Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. Information Fusion, 80, 241-265.
7. Himeur, Y., Rimal, B., Tiwary, A., & Amira, A. (2022). Using artificial intelligence and data fusion for environmental monitoring: A review and future perspectives. Information Fusion.
8. Tsanousa, A., Bektsis, E., Kyriakopoulos, C., González, A. G., Leturiondo, U., Gialampoukidis, I., ... & Kompatsiaris, I. (2022). A review of multisensor data fusion solutions in smart manufacturing: Systems and trends. Sensors, 22(5), 1734.
9. Guo, J., Zhou, Y., Zhang, P., Song, B., & Chen, C. (2021). Trust-aware recommendation based on heterogeneous multi-relational graphs fusion. Information Fusion, 74, 87-95.
10. Malchi, S. K., Kallam, S., Al-Turjman, F., & Patan, R. (2021). A trust-based fuzzy neural network for smart data fusion in internet of things. Computers & Electrical Engineering, 89, 106901.
11. Li, B., Wang, E., Shang, Z., Liu, X., Li, Z., Li, B., ... & Song, Y. (2021). Optimize the early warning time of coal and gas outburst by multi-source information fusion method during the tunneling process. Process Safety and Environmental Protection, 149, 839-849.
12. Zhao, H., Yao, Q., Song, Y., Kwok, J. T., & Lee, D. L. (2021). Side information fusion for recommender systems over heterogeneous information network. ACM Transactions on Knowledge Discovery from Data (TKDD), 15(4), 1-32.
13. Ahmadi, M., Taghavirashidizadeh, A., Javaheri, D., Masoumian, A., Ghoushchi, S. J., & Pourasad, Y. (2022). DQRE-SCnet: a novel hybrid approach for selecting users in federated learning with deep-Q-reinforcement learning based on spectral clustering. Journal of King Saud University-Computer and Information Sciences, 34(9), 7445-7458.
14. Banabilah, S., Aloqaily, M., Alsayed, E., Malik, N., & Jararweh, Y. (2022). Federated learning review: Fundamentals, enabling technologies, and future applications. Information processing & management, 59(6), 103061.
15. Ogundokun, R. O., Misra, S., Maskeliunas, R., & Damasevicius, R. (2022). A review on federated learning and machine learning approaches: Categorization, application areas, and blockchain technology. Information, 13(5), 263.
16. Cheng, S., Wu, J., Xiao, Y., & Liu, Y. (2021). Fedgems: Federated learning of larger server models via selective knowledge fusion. arXiv preprint arXiv:2110.11027.
17. Wang, L. E., Wang, Y., Bai, Y., Liu, P., & Li, X. (2021, May). POI recommendation with federated learning and privacy preserving in cross domain recommendation. In IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (pp. 1-6). IEEE.
18. Yu, R., & Li, P. (2021). Toward resource-efficient federated learning in mobile edge computing. IEEE Network, 35(1), 148-155.
19. Rahman, A., Hossain, M. S., Muhammad, G., Kundu, D., Debnath, T., Rahman, M., ... & Band, S. S. (2023). Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. Cluster computing, 26(4), 2271-2311.
20. Nguyen, D. C., Pham, Q. V., Pathirana, P. N., Ding, M., Seneviratne, A., Lin, Z., ... & Hwang, W. J. (2022). Federated learning for smart healthcare: A survey. ACM Computing Surveys (CSUR), 55(3), 1-37.
21. Ma, X., Zhu, J., Lin, Z., Chen, S., & Qin, Y. (2022). A state-of-the-art survey on solving non-IID data in Federated Learning. Future Generation Computer Systems, 135, 244-258.
22. Rahman, A., Hasan, K., Kundu, D., Islam, M. J., Debnath, T., Band, S. S., & Kumar, N. (2023). On the ICN-IoT with federated learning integration of communication: Concepts, security-privacy issues, applications, and future perspectives. Future Generation Computer Systems, 138, 61-88.
23. Issa, W., Moustafa, N., Turnbull, B., Sohrabi, N., & Tari, Z. (2023). Blockchain-based federated learning for securing internet of things: A comprehensive survey. ACM Computing Surveys, 55(9), 1-43.
24. Wang, T., Liu, Y., Zheng, X., Dai, H. N., Jia, W., & Xie, M. (2021). Edge-based communication optimization for distributed federated learning. IEEE Transactions on Network Science and Engineering, 9(4), 2015-2024.
25. Ahmed, U., Srivastava, G., & Lin, J. C. W. (2022). Reliable customer analysis using federated learning and exploring deep-attention edge intelligence. Future Generation Computer Systems, 127, 70-79.
26. Tan, Y., Long, G., Ma, J., Liu, L., Zhou, T., & Jiang, J. (2022). Federated learning from pre-trained models: A contrastive learning approach. Advances in Neural Information Processing Systems, 35, 19332-19344.

27. Zhang, J., Chen, C., Li, B., Lyu, L., Wu, S., Ding, S., ... & Wu, C. (2022). Dense: Data-free one-shot federated learning. Advances in Neural Information Processing Systems, 35, 21414-21428.
28. Javed, A. R., Hassan, M. A., Shahzad, F., Ahmed, W., Singh, S., Baker, T., & Gadekallu, T. R. (2022). Integration of blockchain technology and federated learning in vehicular (iot) networks: A comprehensive survey. Sensors, 22(12), 4394.
29. Wu, C., Wu, F., Lyu, L., Huang, Y., & Xie, X. (2022). Communication-efficient federated learning via knowledge distillation. Nature communications, 13(1), 2032.
30. Qi, Y., Hossain, M. S., Nie, J., & Li, X. (2021). Privacy-preserving blockchain-based federated learning for traffic flow prediction. Future Generation Computer Systems, 117, 328-337.
31. Jia, B., Zhang, X., Liu, J., Zhang, Y., Huang, K., & Liang, Y. (2021). Blockchain-enabled federated learning data protection aggregation scheme with differential privacy and homomorphic encryption in IIoT. IEEE Transactions on Industrial Informatics, 18(6), 4049-4058.