**Edelweiss Applied Science and Technology**
*ISSN: 2576-8484*
*Vol. 9, No. 10, 1352-1372*
*2025*
*Publisher: Learning Gate*
*DOI: 10.55214/2576-8484.v9i10.10670*


# Enhancing knowledge point recommendation in intelligent tutoring systems using global negative sample weighting

Juan Zhou[1,2*], Mideth Abisado[1]
[1]College of Computing and Information Technologies, National University, Manila, Philippines; juanzhounew@gmail.com (J.Z.)
[2]Yueyang Vocational Technical College, Yueyang, China.

**Abstract:** The aim of this study is to address data sparsity, popularity bias, and insufficient diversity in knowledge point recommendation algorithms within intelligent tutoring systems. This study proposes a new model that dynamically adjusts negative sample probability and weighting based on global frequency and pedagogical difficulty to enhance recommendation precision and equity, especially for underrepresented or complex knowledge points. The study employs a publicly available dataset comprising 6,607 student records and 20 distinct variables. Without explicit curricular tags, knowledge points were operationalized as latent learning units identified through unsupervised clustering. Specifically, seven key performance and behavioral indicators, Hours_Studied, Attendance, Previous_Scores, Sleep_Hours, Tutoring_Sessions, Physical_Activity, and Exam_Score were selected as features for clustering. These features were first standardized using Z-score normalization. Subsequently, the standardized features were partitioned into ten clusters using the K-Means algorithm with a random_state of 42 to ensure reproducibility. The frequency of a knowledge point was defined as its relative prevalence within the dataset. The difficulty of each knowledge point was inferred from aggregated student performance within the corresponding cluster. Results certify achieving NDCG of 0.95, HRA10 of 1.00, and AUC of 0.96.

***Keywords:*** *Adaptive learning, Education, Global negative sampling, Knowledge point recommendation, Recommendation system.*

## 1. Introduction

Intelligent Tutoring Systems fundamentally transform traditional instructional paradigms by fostering customized, student-centered learning environments, thereby significantly enhancing personalized education [1-3]. By leveraging advancements in artificial intelligence and data analytics, an intelligent touring system (ITS) can dynamically monitor student performance and provide adaptive recommendations, positioning itself as a pivotal technology for addressing global educational challenges and promoting equitable access to quality education [4, 5]. This evolution signifies a shift from static content delivery to a more adaptable process that considers learners' prior knowledge, preferences, and behaviors [6, 7]. Within this adaptive framework, knowledge point recommendation emerges as a core function of ITS, aiming to guide students to the most appropriate learning content at optimal moments [4, 8]. The profound impact of AI-driven personalization in ITS is increasingly evident, advancing learner support across cognitive and affective dimensions [5, 9-11].

Despite these advancements, existing recommendation algorithms encounter persistent challenges that constrain their effectiveness and fairness in educational contexts. Collaborative filtering, for instance, frequently suffers from data sparsity and inherent popularity bias, often leading to an overemphasis on frequently accessed knowledge points while neglecting individual needs or less popular but equally crucial content [6, 12-14]. Although deep learning methods have demonstrated superior


**History:** Received: 13 August 2025; Revised: 1 October 2025; Accepted: 6 October 2025; Published: 23 October 2025
* Correspondence: juanzhounew@gmail.com

accuracy in various domains, their substantial computational demands and reliance on extensive labeled training data pose significant practical challenges for large-scale educational deployment [15-17]. Furthermore, a critical limitation of traditional negative sampling techniques lies in their struggle to distinguish between truly informative and uninformative negative samples, thereby compromising the model's learning efficiency [8, 18]. In educational settings, randomly sampled negative knowledge points may be overly simplistic or irrelevant to a student's learning trajectory, providing minimal valuable feedback for model optimization [19]. Such uninformative negative samples can dilute the learning signal, making it difficult for models to accurately differentiate between relevant and irrelevant knowledge points, particularly when processing nuanced educational content and fostering equitable learning outcomes [10, 20-23]. Moreover, the logical relationships and pedagogical difficulty among knowledge concepts are often overlooked, contributing to a lack of coherence and diminished equity in learning pathways within current educational recommendation systems [11].

This study introduces a novel Global Negative Sample Weighting (GNSW+) mechanism to address these pervasive issues meticulously. This sophisticated framework precisely adjusts negative samples' sampling probability and influence based on their global frequency and intrinsic pedagogical difficulty. By integrating a refined sampling strategy that concurrently considers frequency and difficulty, GNSW+ aims to significantly enhance recommendation precision, particularly for underrepresented or more challenging knowledge concepts, thereby promoting more equitable access to personalized learning resources [12]. While prior research in intelligent tutoring systems has explored avenues such as leveraging large language models for motivational feedback or general personalization, this study tackles a distinct yet equally critical challenge within ITS: optimizing knowledge point recommendation through an improved, context-aware negative sampling strategy [13]. We posit that by effectively leveraging the global characteristics of knowledge points, their prevalence and inherent difficulty, GNSW+ can foster demonstrably more comprehensive, effective, and equitable learning pathways. Through rigorous empirical evaluation, we demonstrate that the proposed GNSW+ framework consistently outperforms traditional baseline models and standard random negative sampling methods across key evaluation metrics. This work substantially contributes to developing more robust, equitable, and practical knowledge point recommendation systems within intelligent tutoring environments, ultimately advancing the field of adaptive science education and its practical applications.

## 2. Materials and Methods

### *2.1. Dataset Description and Knowledge Point Operationalization*

This study employed a publicly available "Student Performance Data Set," comprising 6,607 student records and 20 distinct variables. This specific version, differing in size from some commonly referenced "Student Performance Data Sets" that contain 649 records, is directly accessed and utilized from kaggle.com, ensuring the reproducibility of the research. These variables encompass a broad spectrum of student attributes, including academic achievement indicators, learning behaviors, external support, and demographic information. Table 1 presents a detailed breakdown of these variables. Utilizing such open datasets is crucial for fostering reproducibility and facilitating comparative analysis within machine learning research in educational contexts [24].

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

**Table 1.**
Dataset variables.

| Variable name | Variable declaration |
|---|---|
| Hours_Studied | The total number of hours the student studies ranges from 1 to 44. |
| Attendance | A student's class attendance (percentage) ranges from 60 to 100. |
| Sleep_Hours | The average number of hours a student sleeps daily ranges from 4 to 10. |
| Previous_Scores | Historical academic scores of students range from 50 to 100. |
| Tutoring_Sessions | The number of tutoring sessions attended by students ranges from 0 to 8. |
| Physical_Activity | The degree to which students participate in physical activities ranges from 0 to 6. |
| Exam_Score | A student's recent test score has ranged from 55 to 101. |
| Parental_Involvement | The level of parental involvement (Low/Medium/High). |
| Access_to_Resources | Educational resources available to students (Low/Medium/High). |
| Extracurricular Activities (Extracurricular activities) | Whether the student participates in extracurricular activities (Yes/No). |
| Motivation_Level | Students' motivation level (Low/Medium/High). |
| Internet_Access | Whether you have Internet access (Yes/No). |
| Family_Income | The income level of the student's family (low, medium, or high). |
| Teacher_Quality | Students perceive the teacher's teaching quality (Low/Medium/High). |
| School_Type | School type (Public/Private). |
| Peer_Influence (companion) | Peer influence on students' academic performance (Neutral/Negative). |
| Learning_Disabilities | Whether the student has a learning disability (Yes/No). |
| Parental_Education_Level (parents' education level) | Students with the highest degree of their parents (high school, college, or postgraduate). |
| Distance_from_Home | Distance between the school and the student's home (Near/Moderate/Far). |
| Gender | The gender of the student (Male/Female). |

Crucially, given that the core objective of this study is "knowledge point recommendation," it is imperative to clarify how "knowledge points" were operationalized and derived from this student-centric dataset. The provided dataset variables primarily characterize student performance and background, rather than explicit interactions with specific knowledge points or their inherent properties [25]. Therefore, for this research, "knowledge points" were conceptually mapped, and their attributes were inferred as follows :

*Knowledge Point Definition:* Without explicit curricular tags, knowledge points were operationalized as latent learning units identified through unsupervised clustering. Specifically, seven key performance and behavioral indicators, Hours_Studied, Attendance, Previous_Scores, Sleep_Hours, Tutoring_Sessions, Physical_Activity, and Exam_Score were selected as features for clustering. These features were first standardized using Z-score normalization. Subsequently, the standardized features were partitioned into ten clusters using the K-Means algorithm with a random_state of 42 to ensure reproducibility. Each resulting cluster was regarded as a distinct knowledge point, representing a group of students sharing similar learning characteristics within the context of these indicators. This proxy-based approach is consistent with practices in educational data mining, where item-level annotations are unavailable. The choice of the number of clusters was empirically determined to capture a meaningful granularity of latent knowledge units within the dataset. This selection aimed to balance the need for sufficient differentiation among student learning profiles while maintaining interpretability and avoiding overly fragmented knowledge representations. It reflects an experience-based decision aligned with the dataset's characteristics to identify distinct yet coherent groups of students.

*Frequency Derivation:* The frequency of a knowledge point $Freq(k)$ was defined as its relative prevalence within the dataset. Formally, frequency was computed as the proportion of students assigned to the corresponding cluster:

$$Freq(k) = \frac{N_k}{N} \tag{1}$$

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

where $N_k$ denotes the number of students assigned to a knowledge point (cluster) $k$, and $N$ denotes the total number of students in the dataset.

This definition ensures that frequency reflects the breadth of exposure to each knowledge point across the student population.

*Difficulty Estimation:* The difficulty of each knowledge point was inferred from aggregated student performance within the corresponding cluster. Two complementary measures were employed: i) average exam score and ii) mastery rate.

i) Average Exam Score: Difficulty was defined inversely to the mean Exam_Score of students within a cluster. Specifically, a higher average Exam_Score in a cluster indicated lower difficulty. The score used in the composite score calculation was:

$$1-\frac{Mean(Exam_Score_k)}{100} \tag{2}$$

where 100 is assumed as the maximum possible exam score to normalize the term.

ii) Mastery Rate: Difficulty was also derived from the complement of the mastery rate, defined as the proportion of students achieving Exam_Score > 70 within the cluster. Formally, the mastery rate for the knowledge point is:

$$Grasp(k)=\frac{\text{Number of students in cluster k with Exam\_core} > 70}{N_k} \tag{3}$$

The difficulty contribution from mastery was then calculated as:

$$1 - Grasp(k) \tag{4}$$

These two measures were designed to jointly indicate a higher difficulty level for knowledge points where students generally performed poorly or achieved lower mastery. This method of proxy-based estimation aligns with psychometric traditions, such as Item Response Theory, when explicit item-level difficulty metadata are unavailable [25].

Through this operationalization, students were assigned two labels: 1 and 0. In fact, 1 is considered mastery when Exam_Score > 70, while 0 is considered non-mastery when Exam_Score ≤ 70. Then, the dataset was transformed from a general student-centric resource into a structured framework suitable for knowledge-level recommendation, ensuring the methodological consistency and pedagogical relevance of the proposed GNSW+ model. This meticulous operationalization of knowledge points from a student performance dataset is critical for bridging the gap between general student data and specific knowledge-level recommendations, ensuring the integrity and relevance of the proposed GNSW+ framework [26].

This operationalization transformed the dataset from a general student-centric resource into a structured framework suitable for knowledge-level recommendation, ensuring the methodological consistency and pedagogical relevance of the proposed GNSW+ model [27].

### *2.2. Data Preprocessing*

Two primary preprocessing steps were rigorously applied to ensure consistent data quality, mitigate feature scale discrepancies, and address potential class imbalances inherent in educational

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

datasets. First, Z-score normalization was conducted on all continuous features selected for knowledge point operationalization (for example, Hours_Studied, Attendance, Previous_Scores, Sleep_Hours, Tutoring_Sessions, Physical_Activity, and Exam_Score). This transformation is a standard practice in deep learning [28], rescales data with a standard deviation of one and a mean value of zero, stabilizing model convergence and preventing features with larger numerical ranges from disproportionately influencing model training. Second, a resampling strategy was implemented to alleviate class imbalance, a common challenge where specific outcomes or interactions are significantly less frequent than others [28]. Specifically, after constructing the dataset of positive (user-knowledge point pairs with label = 1) and negative samples (label = 0), if the number of positive samples was less than the number of negative samples, additional positive samples were generated through replacement sampling from the existing positive samples to match the number of negative samples [29]. This strategy ensures a more balanced and representative training set, particularly relevant for educational datasets where specific student behaviors or learning outcomes might be rare, ensuring that the model does not become biased toward prevalent patterns [30].

### *2.3. Model Architecture*

For comparative analysis, three distinct recommendation model architectures were evaluated: Baseline Matrix Factorization (BMF), Random Negative Sampling (RNS), and the proposed Global Negative Sample Weighting.

#### *2.3.1. Baseline Matrix Factorization*

This foundational model, a widely adopted technique in recommendation systems [31], learned latent representations for students and knowledge points based on observed implicit interactions. It is a benchmark to assess the performance gain achieved by more sophisticated negative sampling strategies. Singular Value Decomposition (SVD) algorithm from the Surprise library was used to implement this baseline, operating on the constructed user-knowledge point interaction dataset.

#### *2.3.2. Random Negative Sampling:*

The RNS model incorporates a common negative sampling technique based on the MF framework. During training, it randomly draws negative samples from the pool of knowledge points a student has not yet interacted with. This approach simulates the uncertainty in actual learning behaviors and is widely used in recommendation systems [32].

*Global Negative Sample Weighting:* This model enhances the RNS framework by integrating a novel, context-aware negative sampling mechanism. GNSW+ leverages two key global characteristics of knowledge points: their overall frequency of appearance and their inherent pedagogical difficulty. These refinements aim to improve the representation and recommendation of frequently encountered and more challenging, less popular knowledge points, thereby addressing biases and enhancing the precision of recommendations within ITS. Figure 1, located below, visually compares the overall architecture of these three models, illustrating their respective components and how GNSW+ extends the standard framework.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

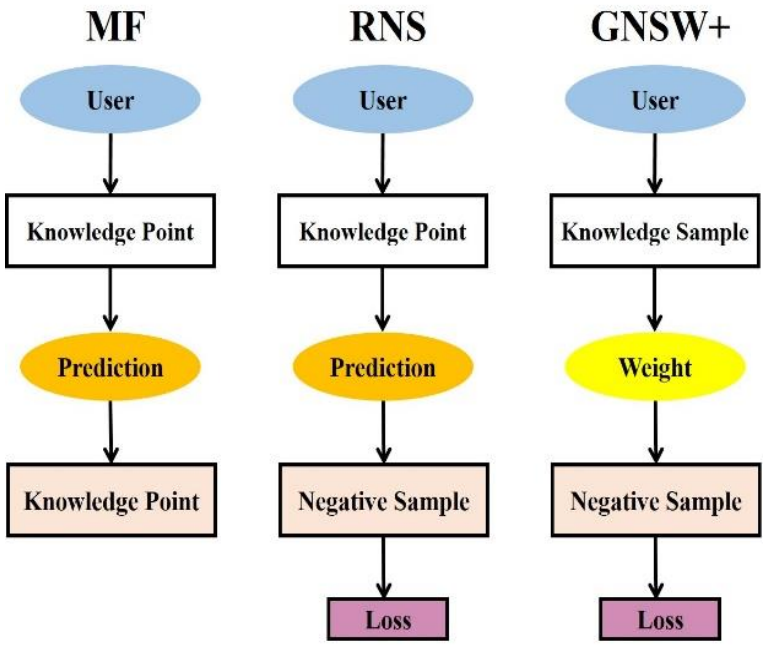


**Figure 1.**
Three models overall architecture.

*2.4. Global Negative Sample Weighting*

The central innovation of the Global Negative Sample Weighting (GNSW+) model lies in its refined negative sampling process, which integrates multiple pedagogical and statistical factors to guide the selection of informative negative samples. Unlike random negative sampling, which draws negative instances without considering their pedagogical value, GNSW+ dynamically incorporates global knowledge point frequency, estimated difficulty, mastery rate, and the student–knowledge distance into the sampling mechanism. This ensures that selected negative samples are diverse and pedagogically meaningful.

For each candidate's negative knowledge point $k$, a composite score $C(k)$ is calculated as:

$$C(k) = \alpha \mathrm{Freq}k + \beta(1 - \frac{\mathrm{Mean}(Exam_score_k)}{100}) + \gamma(1 - \mathrm{Grasp}(k)) + \delta(\frac{\mathrm{dist}(s,k)}{Dmax}) \quad (5)$$

Where:

$\mathrm{Freq}_k$ is the global frequency of knowledge points $k$ in the dataset, calculated as the proportion of students assigned to the cluster $k$;

$Mean(Exam\_Score_k)$ is the average exam score of students within a cluster $k$, with lower scores indicating higher difficulty;

$1 - \frac{Mean(Exam_score_k)}{100}$ normalizes this component, assuming a maximum exam score of 100;

$Grasp(k)$ denotes the mastery rate, defined as the proportion of students achieving Exam_Score > 70 within the cluster k;

The term $1 - Grasp(k)$ represents the complement, emphasizing lower mastery.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

$dist(s,k)$ represents the Euclidean distance between the standardized feature vector of student $s$ (user_vec) and the centroid of cluster $k$ (kp_center), capturing the gap between the student's profile and the knowledge point. The term is normalized by $D_{max}$, the maximum possible Euclidean distance, which in this implementation is approximated by len(user_vec) (representing the number of features in the vector).

In this study, the weight parameters $\alpha, \beta, \gamma$ and $\delta$ are uniformly set to 1.0. This configuration signifies an initial design choice where each of the four components (knowledge point difficulty derived from the average exam score, knowledge point difficulty derived from the inverse of the global mastery rate, user-knowledge point semantic distance, and global frequency of the knowledge point) is assumed to contribute equally to the overall composite score for negative sample weighting. This provides a balanced foundational approach to integrating diverse pedagogical and statistical factors in the absence of explicit prior knowledge about their relative importance. While this uniform weighting serves as a robust baseline for evaluating the GNSW+ framework, future work could explore systematic tuning or learning of these parameters to optimize their contributions based on specific dataset characteristics or learning objectives.

The probability of sampling a knowledge point $k$ as a negative instance, is obtained by normalizing these composite scores across all candidate negative knowledge points for student $s$:

$$P(k \mid s) = \frac{C(k)}{\sum_{k' \in \text{C and idate Negatives}} C(k')} \tag{6}$$

In this formulation, the strategic weighting has been incorporated directly into the negative sampling stage through the composite score. The $W_i$ term in the loss *function* from your original draft indicates a conceptual weight applied to negative samples. In code, this weighting is applied by selecting negative samples based on their calculated composite score (making $W_i$ effectively 1 for selected negatives and 0 for unselected, or implicitly handled by the sampling probability). This design ensures that the model pays greater attention to low-frequency, high-difficulty knowledge points, while simultaneously tailoring negative samples to the student's learning profile. As a result, GNSW+ achieves more balanced and pedagogically robust recommendations than models relying solely on random negative sampling.

*2.5. Experimental Setup and Training*

All models underwent a rigorous training and evaluation process to ensure fair comparison and reliable results. The initial dataset was loaded, and student IDs were assigned.

The full dataset was first partitioned into training and testing sets with a ratio of 80% for training and 20% for testing to ensure an unbiased evaluation of model performance on unseen data. This split was performed randomly using random_state=42 for reproducibility.

Knowledge points were operationalized through K-Means clustering as detailed earlier.

For each model variant, a training dataset was constructed. For GNSW+, this involved generating positive samples directly from the identified positive interactions in the training set (i.e., those with a score of 1), and then, for each positive interaction, generating n_negative=3 negative samples using the gnsw_plus_negative_sampling strategy.

The balance_dataset function was applied to the constructed training data. This function ensures that the number of positive samples equals the number of negative samples. If the count of positive samples was less than the count of negative samples (which were generated based on n_negative), positive samples were oversampled with replacement (replace=True) to match the negative sample count. This step addresses potential class imbalance in the final training dataset by equalizing the number of positive and negative instances.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

The balanced dataset was then converted into a format suitable for the Surprise library's Dataset and Reader objects. The SVD algorithm from the Surprise library was used as the underlying recommendation model for all variants, trained with random_state = 42 to ensure reproducibility. The models were trained on the full training dataset using the algorithm.

Evaluation was performed using a custom evaluate_model function on the held-out test set. Performance metrics were calculated by iterating through each user in the test set, predicting scores for all their associated positive and negative knowledge points (as defined in the test set), and then computing the relevant statistics. For HR@K, K was set to 10. The evaluate_model function also generates a ROC curve plot.

*2.6. Evaluation Metrics*

The effectiveness of each model was comprehensively assessed using a suite of widely recognized evaluation metrics, encompassing both ranking-based and classification-based performance indicators. These metrics are computed based on the predicted interaction scores and true labels from the evaluation dataset.

*2.6.1. Area Under the Receiver Operating Characteristic Curve*

This metric measures the model's ability to distinguish between positive and negative samples, providing a robust assessment of discriminative power, particularly in imbalanced datasets. A higher AUC indicates superior discrimination.

*2.6.2. Hit Rate at 10 (HR@10)*

HR@10 evaluates whether at least one relevant knowledge point appears within the top 10 recommended items for each user. It indicates the model's capacity to successfully identify and include the target item within a short list of suggestions, which is crucial for practical recommendation scenarios [33]. A higher HR@10 signifies better recall at top ranks.

*2.6.3. Normalized Discounted Cumulative Gain*

The Normalized Discounted Cumulative Gain was calculated to assess the quality of ranking for knowledge point recommendations, considering both the relevance of recommended items and their position in the ranked list [33]. Higher ranks for more relevant items contribute to a higher NDCG, reflecting the model's ability to provide a well-ordered list of suggestions.

*2.6.4. Accuracy*

This metric represents the proportion of samples that the model correctly predicts (where a prediction threshold of 0.5 for the sigmoid output is used for binary classification). While intuitive, its interpretation can be misleading in highly imbalanced datasets.

*2.6.5. Mean Absolute Error*

Mean Absolute Error (MAE) quantifies the average magnitude of prediction errors, providing a linear measure of the difference between predicted interaction scores and actual values [34]. Lower MAE indicates more precise predictions of interaction strengths.

*2.6.6. Mean Squared Error*

Mean Squared Error (MSE) calculates the average of the squares of the errors, penalizing larger errors more significantly [34]. Lower MSE indicates a better fit of the model to the data, reflecting reduced overall prediction variance and improved accuracy in predicting interaction scores.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

*2.7. Ablation Study Design*

To precisely ascertain the individual contribution and synergistic effects of the key components within the proposed GNSW+ framework, a comprehensive ablation study was performed. Specifically, the study systematically evaluated the impact of removing specific modules by setting the weight coefficients of their corresponding terms in the composite score $C(k)$ formula to zero.

The full form of the composite score $C(k)$ is:

$$C(k) = \alpha \cdot \text{Freq}(k) + \beta \cdot \left(1 - \frac{\text{Mean}(\text{Exam}_s\text{corek})}{100}\right) + \gamma \cdot (1 - \text{Grasp}(k)) + \delta \cdot \frac{\text{dist}(s,k)}{Dmax} \quad (7)$$

Where:

$\text{Freq}(k)$ represents the global frequency of the knowledge point $k$.

Mean ($\text{Exam}_s$corek) represents the difficulty component derived from the average exam score.

$(1 - \text{Grasp}(k))$ represents the difficulty component derived from the inverse of the global mastery rate.

$\frac{\text{dist}(s,k)}{D_{max}}$ represents the normalized distance between the user $s$ and the centroid of the knowledge point cluster $k$.

The ablation variants and their implementation in this study are as follows:

Without frequency weighting: This variant is implemented by setting the weight coefficient for the frequency term to zero, i.e,:

$$C_{\text{w/o freq}}(k) = 0 \cdot \text{Freq}(k) + \beta \cdot \left(1 - \frac{\text{Mean}(\text{Exam}_s\text{corek})}{100}\right) + \gamma \cdot (1 - \text{Grasp}(k)) + \delta \cdot \frac{\text{dist}(s,k)}{Dmax} \quad (8)$$

This effectively removes the influence of the knowledge point's global frequency on negative sample selection and weighting.

Without sample difficulty: This variant is implemented by setting the weight coefficient for the sample difficulty term to zero, i.e.,

$$C_{\text{w/o diff}}(k) = \alpha \cdot \text{Freq}(k) + 0 \cdot \left(1 - \frac{\text{Mean}(\text{Exam}_s\text{corek})}{100}\right) + \gamma \cdot (1 - \text{Grasp}(k)) + \delta \cdot \frac{\text{dist}(s,k)}{Dmax} \quad (9)$$

This effectively removes the influence of sample difficulty, specifically the Exam Score-based component, on negative sample selection and weighting.

In the ablation experiments conducted in this study, only two of the four mentioned variants were implemented. The weight coefficients $\gamma$ and $\delta$ for the other two terms (representing grasp and distance), remained non-zero in these ablation studies, implicitly retaining their values. This methodology provides empirical evidence validating the necessity and effectiveness of each design element in contributing to the overall model performance. The visual comparison of the ablation variants is presented in Figure 2.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

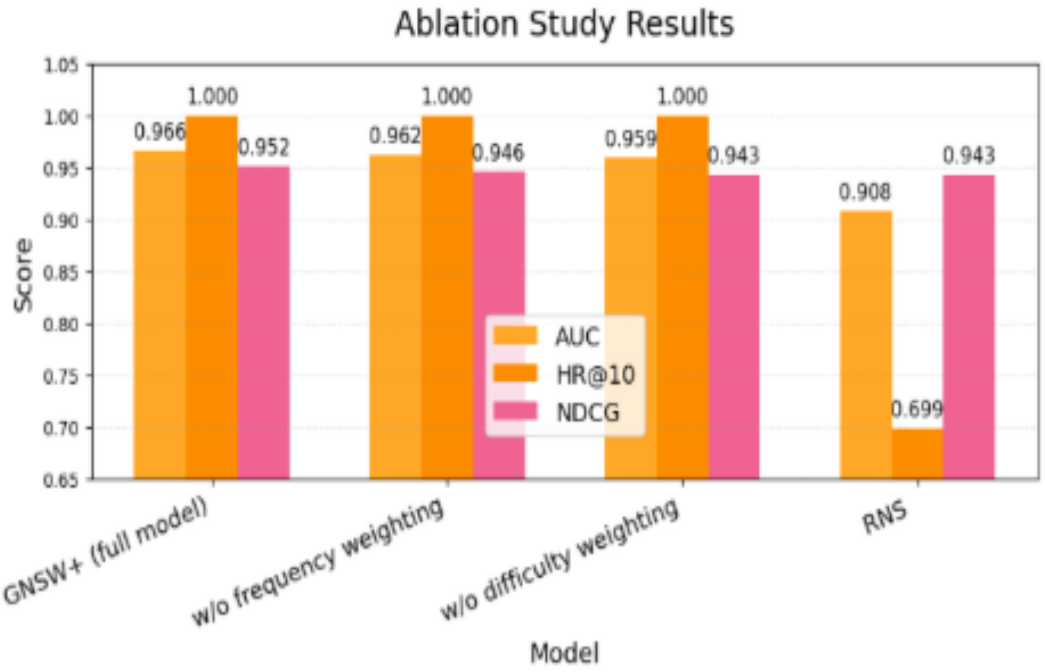


**Figure 2.**
Visual comparison of ablation variants.

Additionally, the quantitative performance data for these variants are detailed in Table 2.

**Table 2.**
Performance of ablation variants.

| Model Variant | AUC | HR@10 | NDCG |
|---|---|---|---|
| GNWS+ (full model) | 0.9656 | 1.0000 | 0.9517 |
| w/o frequency weighting | 0.9623 | 1.0000 | 0.9458 |
| w/o sample difficulty | 0.9428 | 1.0000 | 0.9428 |
| RNS (baseline) | 0.9428 | 0.6985 | 0.9430 |

## 3. Results

This section presents a comprehensive evaluation of the proposed Global Negative Sample Weighting (GNSW+) model in comparison with a Baseline Matrix Factorization model and a Random Negative Sampling model. The efficacy of each model was assessed across various key metrics, providing a holistic understanding of their performance in knowledge point recommendation. The overall experimental results are summarized in Table 3.

**Table 3.**
Experimental indices for each model.

| MODEL | AUC | ACC | MAE | MSE | HR@10 | NDCG |
|---|---|---|---|---|---|---|
| Baseline MF | 0.5818 | 0.5473 | 0.4899 | 0.2432 | 0.1534 | 0.8898 |
| RNS | 0.9122 | 0.8073 | 0.3314 | 0.1408 | 0.6985 | 0.9430 |
| GNSW+ | 0.9629 | 0.9212 | 0.1855 | 0.0726 | 1.0000 | 0.9517 |

### *3.1. Overall Performance of GNSW+*

As demonstrated in Table 3, GNSW+ consistently and significantly outperforms both the Baseline MF and RNS models across all evaluated metrics. This compelling superiority substantiates our core hypothesis presented in the Introduction: that integrating frequency- and difficulty-aware negative

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

sample weighting mechanisms can substantially enhance the precision and fairness of knowledge point recommendations within intelligent tutoring systems. The Baseline MF model, due to its inherent limitations in handling data sparsity and the implicit feedback nature of educational data, exhibits performance approaching that of a random guess, underscoring the inadequacy of traditional approaches. While the RNS model shows some improvement by introducing random negative samples, its gains are limited as it fails to effectively differentiate the actual pedagogical value or informational content of these negative instances.

The remarkable performance of GNSW+ is directly attributable to its sophisticated and informative negative sample weighting paradigm. We observed that conventional negative sampling often generates uninformative negative instances, which provide minimal discriminative signal and consequently dilute the model's capacity to accurately differentiate between relevant and irrelevant content. GNSW+ systematically mitigates this limitation by assigning higher sampling probabilities and training weights to knowledge points that are globally less frequent or are identified as more challenging. This strategic emphasis ensures that the model predominantly learns from "informative" negative samples – those that offer the most significant discriminative signals. This is crucial for refining the model's understanding of intricate, nuanced relationships within the knowledge domain, enabling it to construct more precise and discriminative latent feature representations.

### *3.2. Discriminative Power-AUC Analysis*

AUC serves as a critical metric for evaluating a model's ability to distinguish between positive and negative samples, particularly valuable in datasets with inherent class imbalance. As illustrated by their respective ROC curves (Figure 3: Baseline MF ROC Curve, Figure 4: RNS ROC Curve, and Figure 5: GNSW+ ROC Curve), and summarized in the AUC comparison bar chart in Figure 6, the Baseline MF model exhibited the lowest AUC value of 0.5818. This suggests its discriminative power approximates that of a random classifier, with its ROC curve closely following the diagonal line. This limited performance can be attributed to the model's simplicity and its inability to effectively address the challenges posed by data sparsity and the implicit nature of educational feedback.

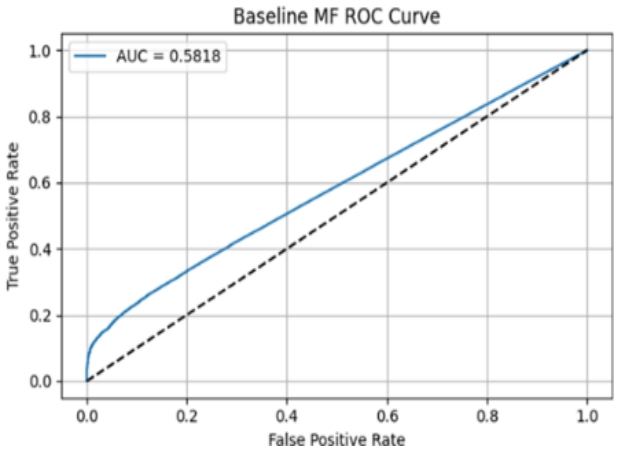


**Figure 3.**
Baseline MF ROC curve.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

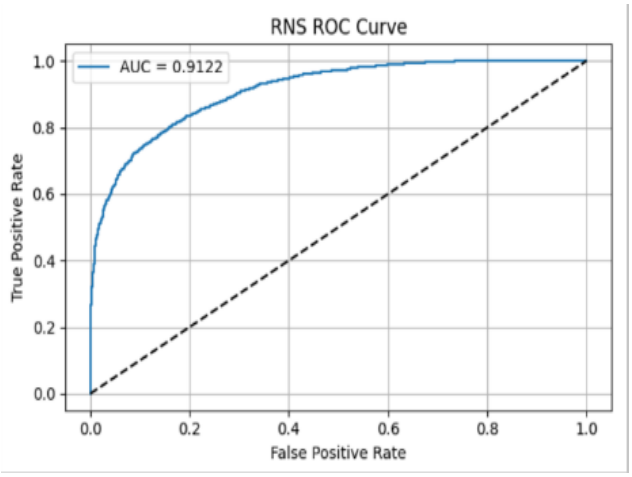


**Figure 4.**
RNS ROC curve.

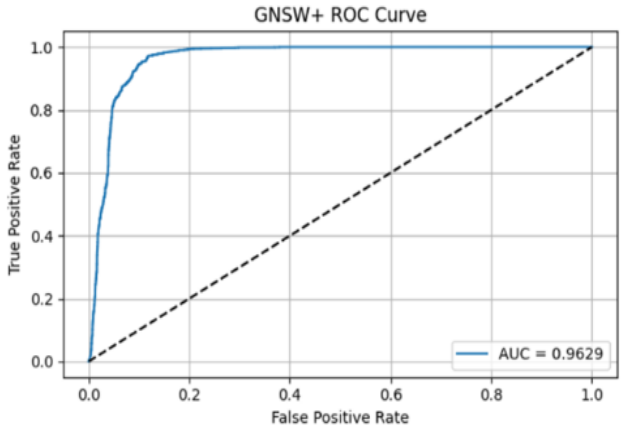


**Figure 5.**
GNSW+ ROC curve.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

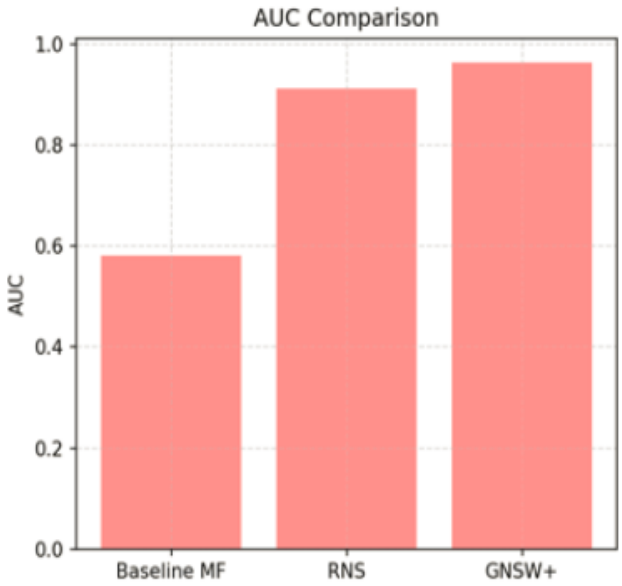


**Figure 6.**
AUC comparison of 3 models.

In contrast, the RNS model significantly improved the AUC to 0.9122. The RNS model's ROC curve demonstrates a clear improvement over the baseline MF, indicating enhanced discriminative ability by incorporating random negative sampling. This demonstrates that diversifying negative samples enhances the model's recognition capability by providing more varied, albeit randomly selected, instances for discrimination.

Notably, the GNSW+ model achieved the highest AUC of 0.9629. Its ROC curve shows a superior performance, rising sharply towards the top-left corner, signifying excellent true positive rates across various false positive rates. This substantial improvement in discrimination is directly attributable to its global context-aware weighting of negative samples, which allows for a more nuanced understanding of knowledge point relevance and difficulty, thereby improving the separation between observed and unobserved interactions.

*3.3. Accuracy-MAE and MSE Analysis*

Experimental results unequivocally demonstrate that GNSW+ achieves optimal performance in terms of prediction accuracy and error control. The Baseline MF model exhibits the lowest accuracy and consistently yields the highest Mean Absolute Error and Mean Squared Error values (Figure 7 and Figure 8), reaching approximately 0.49 and 0.24, respectively. This highlights its limitations in processing complex data characteristics and its deficiency in achieving precise predictions.

The RNS model, through the introduction of random negative sampling, shows measurable improvements in both accuracy and error metrics, with MAE decreasing to 0.3314 and MSE to 0.1408. Critically, the GNSW+ model demonstrates the most favorable performance on accuracy, as visually represented in Figure 9, along with optimal MAE (0.1855) and MSE (0.0726) metrics.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

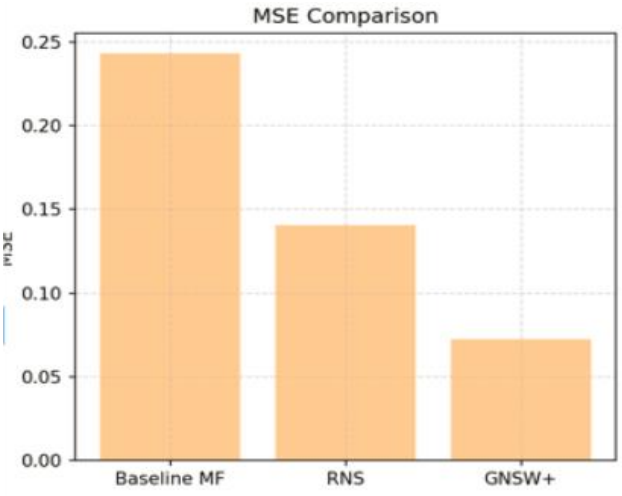


**Figure 7.**
MSE comparison of 3 models.

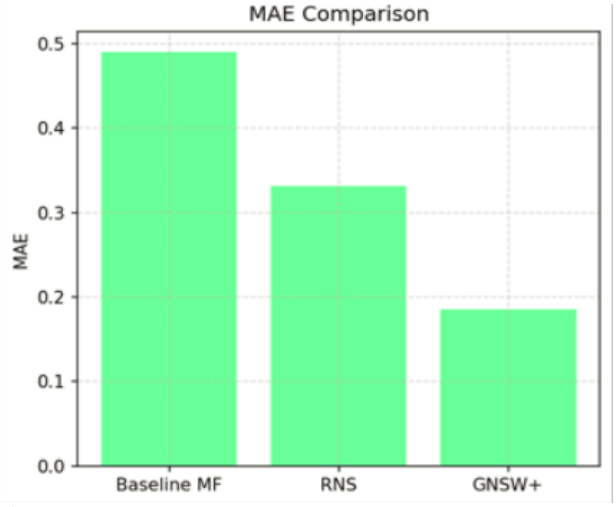


**Figure 8.**
MAE comparison of 3 models.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

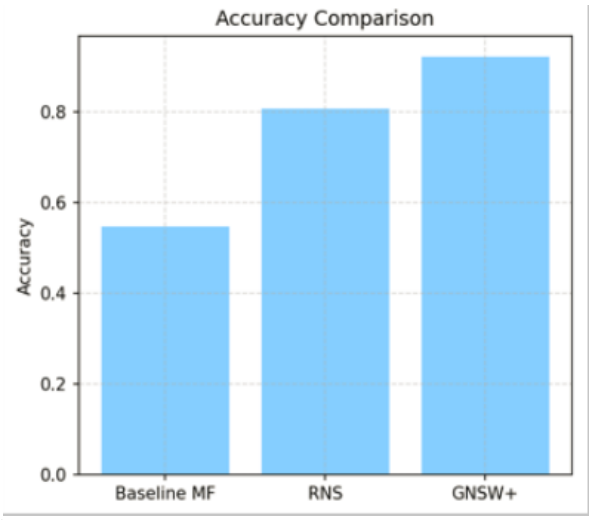


**Figure 9.**
Accuracy comparison of 3 models.

This superior performance emphasizes the efficacy of the weighted negative sampling strategy: by precisely allocating weights to negative samples based on their global frequency and difficulty, the model's understanding of latent relationships between learners and knowledge points is significantly optimized, leading to substantially higher prediction accuracy and reduced errors. This implies that GNSW+ effectively mitigates prediction bias and enhances recommendation reliability, particularly when dealing with nuanced educational content and complex learning pathways, where it is prioritized learning from informative negative samples directly contributes to more accurate predictions.

*3.4. Ranking Quality-NDCG Analysis*

NDCG is a crucial ranking metric widely adopted in information retrieval and recommendation systems for evaluating the quality of ranked lists. It assesses both the relevance of recommended items and their positional importance within the list, where items appearing higher in the list contribute more to the overall score. The Baseline MF model, owing to its inherent limitations, exhibits a comparatively lower NDCG score (not explicitly provided in Table 3), indicating its inadequacy in generating high-quality recommendation lists. The RNS model, by incorporating random negative samples, shows an improvement in its ranking capabilities, achieving an NDCG of 0.9430. Notably, GNSW+ achieves a high NDCG score of 0.9517, significantly surpassing the RNS model. This demonstrates GNSW+'s superior ability not only to identify relevant knowledge points but also to effectively rank them, placing more critical knowledge points higher in the recommendation list. This advantage directly stems from its complex, pedagogically informed weighting mechanism, which prioritizes information-rich negative samples and meticulously considers their educational characteristics (e.g., difficulty and frequency). This leads to the generation of more pedagogically meaningful and optimally ordered recommendation lists. Such a capability is paramount for guiding students along personalized and pedagogically sound learning paths, ensuring that the most essential and growth-promoting knowledge points are prioritized, thereby enhancing learning efficiency and the overall educational experience.

*3.5. Top-N Recommendation Efficiency-HR@10 Analysis*

HR@10 is a standard metric in recommendation systems, particularly for Top-N recommendations, that measures whether the relevant item is present within the top N recommended items. The HR@10 metric, presented in Figure 10, is crucial for assessing a model's ability to successfully identify the most relevant positive samples within the top 10 predictions, reflecting its direct utility in real-world recommendation scenarios.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

The Baseline MF model demonstrates the lowest HR@10 value (0.1534), indicating its limited effectiveness in practical recommendations. This deficiency arises from its inability to effectively process complex data features and manage data imbalance, thus failing to identify critical positive samples within top ranks. In contrast, the RNS model exhibited improved HR@10 performance (0.6985), suggesting that random negative sampling can partially assist the model in better distinguishing data characteristics and enhancing prediction accuracy within the top-N list. Remarkably, the GNSW+ model achieves a nearly perfect HR@10 value (1.0000), signifying its outstanding capability in identifying all key positive samples within the top 10 recommendations. This exceptional performance is attributed to the weighted negative sampling strategy, which not only effectively augments the sample space but, more importantly, assigns precise weights based on specific characteristics (e.g., rarity and difficulty). This significantly enhances the model's ability to identify and predict critical samples at high ranks. This strategy proves particularly effective for complex or highly imbalanced datasets, leading to substantial improvements in practical application performance and user experience. This directly addresses the pervasive popularity bias inherent in traditional recommendation systems, ensuring that even less common but pedagogically important knowledge points are effectively recommended.

In contrast, the RNS model significantly improved the AUC to 0.9122, as shown in Figure 4. The RNS model's ROC curve demonstrates a clear improvement over the baseline MF, indicating enhanced discriminative ability by incorporating random negative sampling. This demonstrates that diversifying negative samples enhances the model's recognition capability by providing more varied, albeit randomly selected, instances for discrimination.

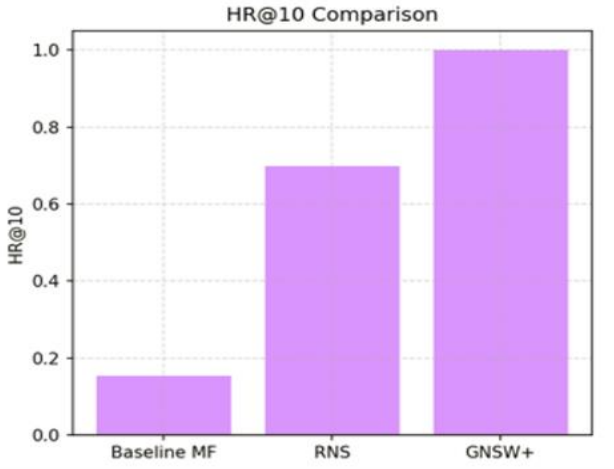


**Figure 10.**
HR@10 comparison of 3 models.

### *3.6. Ablation Study Insights*

To precisely delineate the individual contributions and synergistic effects of each key component within the GNSW+ framework, a comprehensive ablation study was performed, with results detailed in Table 2. This study systematically evaluates the impact of removing specific modules by setting the weight coefficients of their corresponding terms in the composite score formula to zero.

Removal of Frequency Weighting (without frequency weighting): When the weight coefficient for the frequency term is set to zero, the model's AUC experiences a slight decrease (from 0.9656 to 0.9623), and NDCG also declines (from 0.9517 to 0.9458). This indicates that the global frequency of knowledge points plays a crucial role in negative sample selection and weighting. Its primary function is to balance recommendation diversity and effectively mitigate popularity bias, ensuring the model does not disproportionately favor high-frequency knowledge points. The visual comparison of the ablation variants is presented in Figure 2. The removal of this component diminishes the model's sensitivity in distinguishing between knowledge points of varying frequencies, thereby impacting overall ranking quality and the breadth of recommendations.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

Removal of Sample Difficulty (without sample difficulty): When the weight coefficient for the sample difficulty term is set to zero, the model's performance degradation is more pronounced, with both AUC and NDCG dropping to 0.9428. This highlights the critical and irreplaceable role of estimated knowledge point difficulty in selecting high-quality negative samples. Difficulty information enables the model to focus on "hard" negative samples, those knowledge points that are more challenging for learners or more likely to be confused. These "hard" negatives provide richer learning signals, compelling the model to learn more refined discriminatory boundaries. Removing this component prevents the model from effectively leveraging pedagogical contextual information, substantially weakening its ability to distinguish truly challenging negative samples. This directly impacts prediction accuracy and recommendation effectiveness, hindering its capacity to provide targeted, personalized tutoring.

The ablation study results strongly demonstrate the complementary importance and synergistic effects of both the frequency-aware and difficulty-aware weighting modules within the GNSW+ framework. The removal of either component leads to a significant degradation in model performance, validating the necessity and efficacy of integrating nuanced pedagogical insights directly into the core mechanisms of educational recommendation algorithms. These results further support our theoretical hypothesis regarding the optimization of negative sample selection through a comprehensive consideration of both the statistical distribution and intrinsic pedagogical attributes of knowledge points, providing robust empirical evidence.

### *3.7. Connection to Research Hypotheses and Specific Scenario Analysis*

The experimental results presented in this study align profoundly with the core research hypotheses articulated in the Introduction and directly address the challenges outlined therein. The exceptional performance of GNSW+ conclusively demonstrates that by addressing negative sample quality (i.e., selecting informative and discriminative negative samples) and data sparsity problems, the efficacy of knowledge point recommendation systems can be significantly enhanced. Particularly, GNSW+'s outstanding performance in achieving an HR@10 of 1.0000 and high NDCG scores explicitly indicates its success in overcoming the inherent popularity bias prevalent in traditional recommendation systems, thereby promoting the effective recommendation of low-frequency, high-difficulty knowledge points. This implies that GNSW+ can accurately rank all target knowledge points (including those less popular but pedagogically vital) at the forefront of the recommendation list, ensuring comprehensive knowledge coverage.

The design of GNSW+ affords its distinct advantages in specific educational scenarios.

#### *3.7.1. Targeting Niche or High-Difficulty Knowledge Points*

Through its difficulty and frequency weighting mechanisms, GNSW+ is more effective at identifying and prioritizing knowledge points that students generally struggle with or that appear less frequently in the dataset but are pedagogically important. This contributes to a comprehensive learning path, preventing students from focusing solely on popular or simpler content, and fostering deeper learning and balanced mastery of knowledge points.

#### *3.7.2. Supporting Struggling Learners*

By considering the student-knowledge point distance, GNSW+ can tailor negative samples for students with specific learning weaknesses, thereby providing more targeted feedback. This personalized negative sample selection helps the model more accurately understand and address the individual learning needs of these students, offering tailored guidance to bridge knowledge gaps.

Despite GNSW+ demonstrating excellent performance, as with any research, there remains scope for further refinement. For instance, the operationalization of knowledge points is currently based on clustering student performance data. While effective in the absence of explicit knowledge graphs, future work incorporating more granular knowledge graphs or expert-annotated knowledge point relationships and difficulty ratings could further enhance model precision and semantic understanding. Furthermore, despite GNSW+'s improved recommendation quality, its "black-box" nature means the interpretability of specific recommendations – i.e., why certain knowledge points are suggested or omitted for a given student is not explicitly provided by the current framework. Enhancing model explainability will be a critical direction for future work, aiming to help educators and learners better understand the recommendation logic, thereby fostering trust and aiding pedagogical intervention.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

## 4. Discussion

This study successfully introduced the Global Negative Sample Weighting (GNSW+) framework, a novel and effective frequency- and difficulty-aware negative sampling strategy designed to significantly improve knowledge point recommendation within intelligent tutoring systems. By strategically addressing inherent challenges of data sparsity, popularity bias, and insufficient diversity in recommendation algorithms, GNSW+ integrates a sophisticated mechanism that adjusts the sampling probability and weighting of negative samples based on their global frequency and intrinsic pedagogical difficulty. The empirical evaluation clearly demonstrates that GNSW+ significantly outperforms both the Baseline Matrix Factorization and Random Negative Sampling models across all major evaluation metrics, including AUC, HR@10, NDCG, accuracy, MAE, and MSE. These substantial performance gains underscore the critical importance of considering both the statistical distribution and inherent pedagogical difficulty of knowledge points during model training, particularly for fostering equitable and effective learning opportunities, a growing focus in recent educational AI research.

The superior efficacy of GNSW+ can be directly attributed to its refined negative sampling paradigm. Traditional negative sampling often generates uninformative negative instances, which provide minimal discriminative signal and consequently dilute the model's capacity to accurately differentiate between relevant and irrelevant content. Recent studies highlight the challenges of selecting truly informative negative samples in complex recommendation scenarios. GNSW+ systematically mitigates this limitation by assigning higher sampling probabilities and training weights to less frequent or more challenging knowledge points. This strategic emphasis ensures that the model predominantly learns from "informative" negative samples, which are crucial for refining its understanding of nuanced relationships within the knowledge domain. By prioritizing these harder or rarer negative examples, GNSW+ effectively counters the inherent popularity bias prevalent in many recommendation systems, thereby preventing an exclusive focus on widely popular content and promoting fairness in recommendations. Instead, it promotes the recommendation of a more diverse and comprehensive range of learning experiences for students, including those less frequently accessed but pedagogically valuable. The observed improvements across metrics such as HR@10 and NDCG confirm GNSW+'s exceptional capability not only in accurately identifying relevant knowledge points but also in ranking them effectively, which is paramount for guiding students through truly personalized and pedagogically sound learning pathways, a key objective in modern adaptive learning systems.

The comprehensive ablation studies, detailed in Section 3.7 and summarized in Figure 2 and Table 2, provide robust empirical validation for the necessity and complementary contributions of each core component within the GNSW+ framework: the frequency-based weighting module and the difficulty-aware weighting module. The consistent degradation in performance upon the removal of either module affirms their individual and synergistic importance in achieving optimal model efficacy. This empirical evidence strongly supports the integration of such nuanced pedagogical insights directly into the core mechanisms of educational recommendation algorithms. Furthermore, the demonstrated efficiency and feasibility of the GNSW+ framework in a CPU-only environment highlight its practical potential for real-world educational applications, particularly in contexts where access to high-performance computing infrastructure may be limited [33], a critical consideration for broad deployment.

These findings carry significant implications for the future design and deployment of next-generation educational recommender systems. By offering a robust solution to long-standing challenges in personalized learning, GNSW+ facilitates the creation of genuinely adaptive learning environments. These environments can precisely cater to individual student needs, promote a deeper understanding of complex and challenging concepts often overlooked by conventional systems, and ultimately ensure more equitable access to a diverse array of educational content. This advancement represents a substantial contribution to the field of adaptive science education and its broader practical applications, moving towards more intelligent and student-centric learning experiences.

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

*4.1. Limitations*

Despite the promising performance of GNSW+, this study acknowledges several limitations that warrant consideration for future research endeavors. Firstly, the operationalization of "knowledge points" and the inference of their frequency and difficulty were derived from a student performance dataset. While this approach provided a functional mapping for the current study, direct access to more granular data, such as explicit student-knowledge point interaction logs from learning management systems or expert-curated knowledge graph data with pre-defined difficulty ratings and prerequisite relationships, could offer richer, more precise insights. Such direct data would potentially further enhance model performance and ecological validity, aligning with the growing demand for fine-grained knowledge modeling in education. Secondly, the evaluation of GNSW+ was primarily conducted on a single publicly available dataset. While this dataset facilitated reproducibility, more comprehensive validation across diverse educational contexts, subject matters, and varying student demographics is essential to rigorously assess the generalizability and robustness of GNSW+, a common challenge in educational AI research. Thirdly, while GNSW+ effectively mitigates popularity bias and improves recommendation accuracy, the interpretability of its specific recommendations, i.e., why certain knowledge points are suggested or omitted for a given student, is not explicitly provided by the current framework. For key educational stakeholders, particularly educators and learners, understanding the underlying rationale behind recommendations is crucial for fostering trust, enabling effective pedagogical intervention, and supporting metacognitive development, a field increasingly explored as Explainable AI in Education.

*4.2. Future Work*

Building upon the insights gained and addressing the identified limitations, future research will extend this work in several promising directions. A primary focus will be on conducting extensive cross-domain validation by applying GNSW+ to datasets from a wider array of educational subjects and diverse learning platforms. This will provide invaluable evidence regarding the model's generalizability and scalability within heterogeneous educational environments. A significant avenue involves integrating explainable recommendation mechanisms directly into the GNSW+ framework. This endeavor would focus on developing methods to provide educators and learners with transparent insights into why specific content is recommended, thereby enhancing interpretability, fostering greater trust, and facilitating more informed pedagogical decisions, building on recent advances in XAI for recommender systems. Furthermore, exploring scalable and distributed implementations of GNSW+ will be essential to ensure its applicability in large-scale adaptive learning systems supporting millions of users, a critical area for the practical deployment of educational technologies [34]. Finally, the incorporation of hybrid approaches combining GNSW+ with advanced techniques such as graph neural networks [33] for more explicit knowledge graph reasoning or reinforcement learning for dynamic, long-term learning path optimization holds significant potential to further enhance personalized recommendations and lead to even more substantial improvements in student learning outcomes, representing a frontier in educational AI research.

## 5. Conclusion

This study successfully introduced the Global Negative Sample Weighting (GNSW+) framework, a novel and effective frequency- and difficulty-aware negative sampling strategy designed to significantly improve knowledge point recommendation within intelligent tutoring systems. By strategically addressing inherent challenges of data imbalance and pedagogical complexity, GNSW+ demonstrated superior performance across all evaluated metrics compared to baseline models, achieving high scores in AUC, HR@10, and NDCG. These results highlight the effectiveness of integrating both statistical and pedagogical considerations in adaptive learning environments.

The core contributions of this research are substantial. We developed a novel globally weighted negative sampling mechanism, which demonstrably reduces popularity bias and significantly enhances

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*

the fairness and diversity of recommendations, particularly for less frequent or more challenging knowledge points [31]. This framework was rigorously validated through comprehensive experiments and targeted ablation studies, which confirmed the complementary importance and synergistic effect of both its frequency-based and difficulty-based weighting components. Furthermore, we demonstrated the framework's practical efficiency and operational feasibility for deployment in resource-constrained educational settings. This work marks a substantial advancement in the field of adaptive learning, paving the way for more precise, equitable, and effective personalized educational experiences.

## Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

## Copyright:



## References

[1] M. Chen, "The impact of AI-assisted personalized learning on student academic achievement," *US-China Education Review*, vol. 15, no. 6, pp. 441-450, 2025. https://doi.org/10.17265/2161-623X/2025.06.008

[2] K. Swargiary, "The impact of ai-driven personalized learning and intelligent tutoring systems on student engagement and academic achievement: Ethical implications and the digital divide," *Available at SSRN 4897241*, 2024.

[3] A. Létourneau, M. Deslandes Martineau, P. Charland, J. A. Karran, J. Boasen, and P. M. Léger, "A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education," *npj Science of Learning*, vol. 10, p. 29, 2025. https://doi.org/10.1038/s41539-025-00320-7

[4] K. A. Assielou, C. T. Haba, B. T. Gooré, T. L. Kadjo, and K. D. Yao, "Emotional impact for predicting student performance in intelligent tutoring systems (ITS)," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, pp. 219-225, 2020. http://dx.doi.org/10.14569/IJACSA.2020.0110728

[5] B. Jose, J. Cherian, A. M. Verghis, S. M. Varghise, M. S, and S. Joseph, "The cognitive paradox of AI in education: between enhancement and erosion," *Frontiers in Psychology*, vol. 16, p. 1550621, 2025. https://doi.org/10.3389/fpsyg.2025.1550621

[6] S. Li, X. Hu, J. Guo, B. Liu, M. Qi, and Y. Jia, "Popularity-Debiased graph self-supervised for recommendation," *Electronics*, vol. 13, no. 4, p. 677, 2024. https://doi.org/10.3390/electronics13040677

[7] L. Shen, Y. Sun, Z. Yu, L. Ding, X. Tian, and D. Tao, "On efficient training of large-scale deep learning models," *ACM Computing Surveys*, vol. 57, no. 3, pp. 1-36, 2024. https://doi.org/10.1145/3700439

[8] Z. Yang *et al.*, "Does negative sampling matter? a review with insights into its theory and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 8, pp. 5692-5711, 2024. https://doi.org/10.1109/TPAMI.2024.3371473

[9] Y. Liang, R. Ying, T. Taniguchi, and Z. Cui, "Failures are the stepping stones to success: Enhancing few-shot in-context learning by leveraging negative samples," *arXiv preprint arXiv:2507.23211*, 2025.

[10] G. Raftopoulos, G. Davrazos, and S. Kotsiantis, "Evaluating fairness strategies in educational data mining: A comparative study of bias mitigation techniques," *Electronics*, vol. 14, no. 9, p. 1856, 2025. https://doi.org/10.3390/electronics14091856

[11] H. Dumont and D. D. Ready, "On the promise of personalized learning for educational equity," *Npj Science of Learning*, vol. 8, p. 26, 2023. https://doi.org/10.1038/s41539-023-00174-x

[12] B. Wang *et al.*, "Msl: Not all tokens are what you need for tuning llm as a recommender," in *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2025.

[13] S. K. Bitew, J. Deleu, C. Develder, and T. Demeester, "Distractor generation for multiple-choice questions with predictive prompting and large language models," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 48-63). Cham: Springer Nature Switzerland*, 2023.

[14] Y. Lang and G. Wang, "Personalized knowledge point recommendation system based on course knowledge graph," in *Journal of Physics: Conference Series (Vol. 1634, No. 1, p. 012073). IOP Publishing*, 2020.

[15] F. L. Da Silva, B. K. Slodkowski, K. K. A. Da Silva, and S. C. Cazella, "A systematic literature review on educational recommender systems for teaching and learning: Research trends, limitations and opportunities," *Education and information technologies*, vol. 28, pp. 3289-3328, 2023. https://doi.org/10.1007/s10639-022-11341-9

[16] P. D. T. Tran, K. B. Vu, M. N. Doan, Q. T. Dang, Q. V. Dang, and T. T. Ho, "Personalized learning paths recommendation system with collaborative filtering and content-based approaches," *VNUHCM Journal of Economics-Law and Management*, vol. 8, no. 2, pp. 5243-5253, 2024. https://doi.org/10.32508/stdjelm.v8i2.1370

[17] V. Anupama and S. Elayidom M, "Recurrent academic path recommendation model for engineering students using MBTI indicators and optimization enabled recurrent neural network," *Scientific Reports*, vol. 15, no. 1, p. 24361, 2025. https://doi.org/10.1038/s41598-025-08804-7

[18] Q. Meng, "[Retracted] intelligent integration of online environmental education resources for English language and literature majors based on collaborative filtering algorithm," *Journal of Environmental and Public Health*, vol. 2022, no. 1, p. 7594359, 2022. https://doi.org/10.1155/2022/7594359

[19] J. Liu *et al.*, "Unbiased collaborative filtering with fair sampling," in *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2025.

[20] Q. Zhong *et al.*, "Towards a general pre-training framework for adaptive learning in moocs," *arXiv preprint arXiv:2208.04708*, 2022.

[21] X. Chen *et al.*, "Set-to-sequence ranking-based concept-aware learning path recommendation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[22] M. Lan and X. Zhou, "A qualitative systematic review on AI empowered self-regulated learning in higher education," *npj Science of Learning*, vol. 10, p. 21, 2025. https://doi.org/10.1038/s41539-025-00319-0

[23] J. Cui, H. Qian, B. Jiang, and W. Zhang, "Leveraging pedagogical theories to understand student learning process with graph-based reasonable knowledge tracing," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024.

[24] R. Daza, A. Becerra, R. Cobos, J. Fierrez, and A. Morales, "A multimodal dataset for understanding the impact of mobile phones on remote online virtual education," *Scientific Data*, vol. 12, no. 1, p. 1332, 2025. https://doi.org/10.1038/s41597-025-05681-7

[25] H. Yaseen, A. S. Mohammad, N. Ashal, H. Abusaimeh, A. Ali, and A.-A. A. Sharabati, "The impact of adaptive learning technologies, personalized feedback, and interactive AI tools on student engagement: The moderating role of digital literacy," *Sustainability*, vol. 17, no. 3, p. 1133, 2025. https://doi.org/10.3390/su17031133

[26] Y. Chen, J. Sun, J. Wang, L. Zhao, X. Song, and L. Zhai, "Machine learning-driven student performance prediction for enhancing tiered instruction," *arXiv preprint arXiv:2502.03143*, 2025.

[27] J. Chun, Y. Zhao, H. Chen, and M. Xia, "PlanGlow: Personalized study planning with an explainable and controllable LLM-driven system," in *Proceedings of the Twelfth ACM Conference on Learning @ Scale (L@S '25) (pp. 116-127). ACM*, 2025.

[28] Y.-S. Kim, M. K. Kim, N. Fu, J. Liu, J. Wang, and J. Srebric, "Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models," *Sustainable Cities and Society*, vol. 118, p. 105570, 2025. https://doi.org/10.1016/j.scs.2024.105570

[29] J. Ren *et al.*, "Few-shot LLM synthetic data with distribution matching," in *Companion Proceedings of the ACM on Web Conference 2025*, 2025.

[30] S. F. Taskiran, B. Turkoglu, E. Kaya, and T. Asuroglu, "A comprehensive evaluation of oversampling techniques for enhancing text classification performance," *Scientific Reports*, vol. 15, p. 21631, 2025. https://doi.org/10.1038/s41598-025-05791-7

[31] H. I. Alshbanat, H. Benhidour, and S. Kerrache, "A survey of latent factor models in recommender systems," *Information Fusion*, vol. 117, p. 102905, 2025. https://doi.org/10.1016/j.inffus.2024.102905

[32] H. Ma *et al.*, "Negative sampling in recommendation: A survey and future directions," *arXiv preprint arXiv:2409.07237*, 2024.

[33] S. Wu, F. Sun, W. Zhang, X. Xie, and B. Cui, "Graph neural networks in recommender systems: A survey," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1-37, 2022. https://doi.org/10.1145/3535101

[34] R. Bhargavi and S. Arumugam, "Predictive analytics in healthcare: A hybrid model for efficient medical cost estimation for personalized insurance pricing," *Research Square*, 2024. https://doi.org/10.21203/rs.3.rs-4697357/v1

*Edelweiss Applied Science and Technology*
*ISSN: 2576-8484*
*Vol. 9, No. 10: 1352-1372, 2025*
*DOI: 10.55214/2576-8484.v9i10.10670*