MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2025, 21(7), em2666
ISSN:1305-8223 (online)
OPEN ACCESS Research Paper https://doi.org/10.29333/ejmste/16604

Data-driven recommendation system for calculus learning using Funk-SVD: Evidence from a mid-scale case study

Chih-Cheng Hsu [1], Che-Yu Hsu [1*]

[1] Research Center for Science and Technology for Learning, National Central University, Taoyuan, TAIWAN

Received 31 March 2025 ▪ Accepted 20 June 2025

Abstract

This study leverages student performance data and the Funk-singular value decomposition (Funk-SVD) model to identify conceptual weaknesses in first-year calculus learning and generate targeted practice recommendations. Rather than relying on error counts or instructor judgment, the model infers individual learning gaps based on predicted success probabilities. Using data from six exams administered to 850 students, the model achieved strong predictive performance, with an F1-score of 0.794. Simulated intervention analysis revealed that the most substantial learning gains occurred among lower-achieving students. Frequently recommended items indicated persistent difficulties with volume integration, curvature, and Riemann sums. These findings underscore the potential of advanced recommendation models to support scalable, personalized learning–grounded in precise, data-informed diagnosis of conceptual weaknesses–thereby enabling more effective instructional support and promoting long-term academic continuity.

Keywords: calculus, Funk-SVD, recommendation systems, data-driven instruction

INTRODUCTION

Calculus represents a foundational subject in higher education, essential not only for mathematics majors but also for students across various specialized disciplines such as engineering, computer science, and the natural sciences. In Taiwan, where the economy is driven by industries such as electrical engineering, computer science, and semiconductors, proficiency in calculus is essential for future professionals. However, the abstract nature and significant conceptual shift from high school mathematics pose substantial challenges.

A unique factor in Taiwan's education system further compounds this issue. Although calculus is introduced in the 12th grade, it is not covered in the primary college entrance examination–the general scholastic ability test. Consequently, many students enter university without a solid calculus foundation, resulting in pronounced academic struggles, particularly among first-year students. Early difficulties with calculus concepts often correlate with course failure, impeding academic progression and essential skill development.

Moreover, the conceptual complexity of calculus demands a deeper understanding of function properties and abstract reasoning. Unlike procedural mathematics emphasized in high school, where speed and algorithmic proficiency are typically emphasized. Calculus was introduced as a broader range of problem-solving strategies and conceptual frameworks. Students often perceive calculus merely as a collection of definitions and formulas, without grasping underlying principles (Cheng, 2007). This disconnect often leads to surface learning approaches, further exacerbating difficulties in mastering the subject.

Recent institutional data further highlights the extent of this issue. At many universities in Taiwan, first-year calculus courses report high rates of failure or course withdrawal, particularly among students from non-mathematics backgrounds. Common difficulties include gaps in fundamental concepts such as limits, derivatives,

A preliminary version of this study was presented at the 8th Eurasian Conference on Educational Innovation, held in Bali, Indonesia, from February 7 to 9, 2025.

✉ waderhsu@gmail.com ✉ cheyuhsu37@gmail.com **(*Correspondence)**

Contribution to the literature

- This study bridges the gap between recommendation technologies and data-driven instructional strategies. Also the results demonstrate the potential of Funk-SVD to enhance learning outcomes in higher education.
- By identifying individual conceptual weaknesses and recommending targeted practice, the system enables instructors to design remedial plans, adapt teaching content, and allocate resources more efficiently.
- Beyond demonstrating the system's predictive accuracy, this study highlights its potential applications for personalized learning and instructional resource management.

and integral applications, as well as challenges in transitioning from procedural computation to conceptual understanding. These struggles not only impact academic performance but also contribute to decreased confidence and motivation, which may have long-term consequences on students' educational trajectories. In fact, persistent failure in foundational subjects like calculus is frequently cited as a key factor influencing students' decisions to transfer to other institutions after their first year–or even to discontinue their university studies entirely.

In light of these challenges, we employ the Funk-singular value decomposition (Funk-SVD) model to analyze data from six exams administered over two semesters. The dataset includes responses from more than 850 students and 138 distinct calculus problems. Unlike traditional diagnostic methods, this approach captures patterns in students' problem-solving behaviors without requiring manual content analysis, thereby enabling fine-grained, data-driven predictions of individual student performance.

Building upon these predictive insights, the system provides tailored practice recommendations designed to strengthen each student's proficiency in calculus. At the same time, it offers instructors valuable feedback to refine their teaching strategies. This dual approach supports both personalized learning and data-driven pedagogical decision-making.

Moreover, the scalability and adaptability of the proposed system suggest potential applications beyond the domain of calculus, contributing to the broader discourse on educational analytics and personalized learning systems.

Accordingly, this study is guided by the following **research questions:**

RQ1. How accurately can the Funk-SVD model predict individual student performance?

RQ2. Can a personalized practice recommendation system effectively enhance students' calculus learning outcomes while supporting instructors in adjusting their teaching strategies?

RQ3. Which groups of students are likely to benefit the most from this system?

RQ4. What are the primary weaknesses in students' calculus problem-solving abilities?

RELATED WORK

Teaching and Learning of Calculus

Mathematics educators generally recognize that the concepts in calculus pose significant challenges for learners (Baker et al., 2000). In the initial learning phase, limits serve as the foundation for understanding derivatives and integrals, making them a primary concept for students to grasp (Sofronas et al., 2011). Understanding limits requires a type of reasoning distinct from algebraic thinking, which many students struggle to develop (Cappetta & Zollman, 2013). Additionally, the use of language and terminology can interfere with students' comprehension of limits. Bressoud et al. (2016) and Larsen et al. (2017), citing Cornu's (1981, 1983, 1991) research, which found that students, influenced by everyday language, often interpret the term "approach" as a process of getting closer to an unattainable value. Consequently, they perceive the "limit" as a boundary that should not be surpassed. When instructors use phrases such as "sufficiently close", "arbitrarily close", or "as close as desired" in the classroom, students frequently struggle to conceptualize the meaning of "close enough" (Keene et al., 2014; Swinyard, 2011; Swinyard & Larsen, 2012). This highlights the crucial role of understanding the formal definition of limits in the development of students' higher-order thinking skills (Kidron, 2020; Parr, 2023; Thompson & Harel, 2021). Zhang (2022) conducted a study on applied mathematics students' learning difficulties in limits and found prevalent misconceptions (Chang, 2021), including an inability to accurately describe both the intuitive meaning and the formal definition of limits, insufficient understanding of the conditions underlying operational rules and notation, and difficulty distinguishing between limit values, function values, and continuity. Furthermore, students encountered challenges when working with limits involving radicals, absolute values, and periodic functions, with most failing to apply the precise definition of limits in proofs. The study ultimately recommended that instruction should reinforce students' conceptual understanding and reasoning abilities regarding limits. In mathematics education, conflicts between language and conceptual understanding are not merely cognitive issues but also concern the precision of instructional language. Thus, when teaching limits and related concepts, instructors must ensure that students correctly interpret these abstract ideas to establish a solid foundation for further learning.

Within the core topics of differential calculus, derivatives represent a fundamental concept that students must comprehend (Aydin & Ubuz, 2014). The three sub-concepts of derivatives include rate of change, graphical representation, and computational methods (Sofronas et al., 2011). While students often manage to execute differentiation procedures, they struggle to understand the meaning of derivatives in terms of rates of change, quantitative reasoning, and the relationships between rate of change, slope, and function behavior (Thompson & Harel, 2021). Dilling and Witzke (2020) emphasized the need to link derivatives to the concept of slope by interpreting derivatives as the slope of the tangent line, aiding students in developing a more meaningful and coherent understanding. Studies have shown that students perform well when dealing with derivatives in kinematics contexts but face greater difficulties interpreting derivatives in non-kinematic contexts. This discrepancy may arise because instructors frequently frame derivative problems in terms of motion; thus, integrating other real-world applications, such as economics, may help students develop a more comprehensive understanding of derivatives and their applications across disciplines (Mkhatshwa, 2024). Additionally, students often struggle with conceptual misunderstandings and procedural deficiencies when learning about inverse functions. Conceptual challenges include failing to recognize the relationships between trigonometric functions and their inverses, forgetting the domain and range of trigonometric functions, and misinterpreting the concept of symmetry. Procedural difficulties involve proving trigonometric identities, graphing inverse functions, and applying inverse function concepts (Ancheta, 2022).

In integral calculus, integration is defined as both the inverse operation of differentiation, which calculates instantaneous rates of change, and as the summation of infinitely many small quantities to determine a whole (Berggren, 2016). However, many students resort to rote memorization of theorems and formulas without grasping the underlying concepts or problem-solving strategies (Idris, 2001). As a result, they fail to comprehend the inverse relationship between integration and differentiation, leading to errors when reverse operations are required. For instance, in integration by parts, students may mistakenly differentiate the integral component instead of applying the correct formula (Angco, 2021). Common difficulties in integration include uncertainty about when to apply integration by parts, confusion in selecting u and d_v, miscalculations during integration or differentiation, misunderstanding substitution methods, and failing to correctly adjust limits or variables after substitution (Ferrer, 2016). In partial fraction decomposition, students may struggle with decomposing rational functions or making errors in the subsequent integration process (Angco, 2021). Research has identified fundamental reasoning skills necessary for constructing definite integrals, including

(1) a dynamic view of limits, where a finite quantity is partitioned into infinitely many or uncountable parts,
(2) a metric perspective on limits, understanding infinitesimally small distances and convergence,
(3) the feasibility of partitioning a finite distance into infinitely many segments,
(4) the existence of infinitely many real numbers within any numerical interval,
(5) the interpretation of a circle as a geometric shape with an infinite number of edges approaching zero in length,
(6) the ability to approximate a curve by partitioning it into rectangular segments, and
(7) the understanding that if the sum of small area elements approaches infinity, the total area error approaches zero (López-Leyton et al., 2024).

Moreover, students commonly make errors in setting integration bounds, determining the correct radius of rotation, computing the volume of hollow solids, and applying volume formulas for solids of revolution (Ting, 2018).

Regarding students’ cognitive learning processes, research suggests that successful calculus learning depends on a strong foundation in prerequisite knowledge such as algebra, trigonometry, and analytic geometry. Many students underestimate the importance of these foundational skills, leading to difficulties in calculus (Angco, 2021). Students often perceive limits, derivatives, and integrals as isolated topics, which hinders their ability to establish conceptual connections (Martin, 2013), thereby making success in calculus more challenging. This highlights that learning calculus is not merely a technical exercise but a process of understanding mathematical structures and modes of reasoning. Consequently, instruction should guide students to build connections among different calculus topics. The abstract nature of calculus can negatively impact students’ attitudes toward mathematics and science (de Vera et al., 2022; Kunwar, 2021), affecting their motivation and learning outcomes. Studies suggest that emphasizing the relevance of calculus in real-world applications, providing ample practice opportunities, and incorporating engaging and challenging calculus problems can enhance students’ interest and confidence in learning calculus (Hammoudi & Grira, 2023).

Furthermore, previous studies have emphasized that a strong foundation in algebra, trigonometry, and analytic geometry is essential for success in entry-level calculus (Ferrini-Mundy & Gaudard, 1992; Hurdle & Mogilski, 2022; Pyzdrowski et al., 2013; Schraeder et al., 2019). These prerequisite domains underpin the development of symbolic manipulation skills and functional reasoning, both of which are critical for calculus learning. Although these topics are formally covered in Taiwan's high school mathematics curriculum, deficiencies in prerequisite knowledge remain evident in students' actual performance (Angco, 2021; Mahadewsing et al., 2024).

To address this gap, the present study adopts a concept-tagged item framework that explicitly integrates both calculus concepts and prerequisite knowledge domains. By applying the Funk-SVD model to student response data, we identify not only performance deficiencies in core calculus areas but also persistent gaps in foundational mathematical understanding. This approach allows the system to generate diagnostic insights that inform targeted instructional planning and support personalized learning trajectories, thereby enhancing both teaching effectiveness and student learning outcomes.

Singular Value Decomposition

SVD is an important matrix decomposition method in linear algebra, widely applied in data compression, dimensionality reduction, signal processing, and machine learning. Billsus and Pazzani (1998) were the first to apply SVD to recommender systems, where it was considered one of the best-performing machine learning algorithms for collaborative filtering (CF) at the time (Zhang, 2022). The primary objective of traditional SVD is to decompose a rating matrix into the product of three matrices (A = UΣVT), thereby enabling dimensionality reduction and imputing missing rating values.

SVD decomposes the entire matrix (e.g., a user-item rating matrix). Suppose there are 1 million users and 100,000 items, forming a 1,000,000 × 100,000 rating matrix where most ratings are missing. When SVD operates on the entire matrix, it generates three matrices, which are often similar in size to or even larger than the original matrix, requiring a significant amount of memory for storage. The resulting matrices tend to be dense (even if the original matrix was sparse), making large-scale sparse matrix computations complex and memory intensive.

To overcome these drawbacks, Funk (2006) proposed Funk-SVD in 2006, which employs an explicit learning process. The so-called "explicit learning process" means that the model has a well-defined objective function (usually a loss function such as mean squared error) and a concrete optimization method (e.g., gradient descent) during training. By continuously adjusting the model parameters to minimize the loss function, the predicted results become closer to the actual data. This method decomposes the rating matrix into the product of two low-dimensional matrices (commonly referred to as P and Q) and learns their parameters by minimizing the loss function on the training data (Zhang, 2022).

Funk-SVD performs well in recommendation tasks mainly because it does not directly rely on surface-level similarities between users or items. Instead, it uncovers deeper preference correlations through latent factor learning. Thus, Funk-SVD is also known as the latent factor model (LFM) (Bi et al., 2019).

Building upon Funk-SVD, Koren (2010) introduced SVD++ in the Netflix Prize competition. SVD++ extends the LFM model by incorporating information from users' historically rated items and considering neighborhood effects. This improvement is an explicit learning enhancement of the item-to-item CF algorithm (Linden et al., 2003), making SVD++ an extension of Funk-SVD that accounts for implicit ratings.

Factorization machines (FM) were later introduced to integrate the advantages of support vector machines and matrix factorization (MF). FM is designed to handle large-scale sparse data–datasets where most values are missing, empty, or zero–and can generalize MF and SVD++ (Rendle, 2010). Unlike traditional SVD-based approaches, FM not only utilizes first-order linear features but also considers second-order (pairwise interaction) features to analyze deeper relationships between variables.

In recent years, deep factorization machines (DeepFM) have emerged as an evolution of the wide and deep model. DeepFM aims to improve click-through rate prediction by modeling both low-order and high-order feature interactions (Zhang, 2022).

In summary, the evolution of SVD-based recommendation systems has progressed from traditional SVD to Funk-SVD (LFM) for improved efficiency and accuracy, then to SVD++ incorporating neighborhood effects, followed by FM with enhanced feature interaction capabilities, and finally to DeepFM, which integrates MF techniques into deep learning frameworks. This development trajectory reflects the ongoing pursuit of higher recommendation quality and the ability to handle increasingly complex data in the field of recommender systems.

Mathematical Model

Funk-SVD assumes that each user can be represented by a set of latent factors, such as preferences for action movies, romance movies, etc. Similarly, each item (e.g., a movie) can also be represented by a set of latent factors describing its characteristics. The rating a user gives to an item can be approximated by taking the inner product of the "user's latent factor vector" and the "item's latent factor vector."

Model Operation

Funk-SVD learns only from known ratings, significantly reducing computational resources. During the training process, the model initializes the latent factor vectors randomly. These vectors are then iteratively adjusted to minimize the error between the predicted ratings and the actual ratings.

Error and Optimization

The model adjusts the latent factor vectors of users and items by calculating the difference between the predicted and actual ratings, known as the "prediction error." This process utilizes a method called stochastic gradient descent, which continuously fine-tunes the parameters based on the error until the overall prediction error is minimized.

To prevent overfitting (where the model becomes too closely fitted to the training data and loses generalization ability), a regularization term is added. This helps control the magnitude of the latent factor vectors, ensuring the model remains stable and maintains predictive accuracy.

Assume $R_{u,i}$ be the actual rating given by user u for item i, p_u be the latent factor vector of user u, and q_i be the latent factor vector of item i. The predicted rating is given by Eq. (1):

$$R_{u,i} = p_u q_i = \sum_{f=1}^{k} p_{u,f} x\, q_{i,f}. \quad (1)$$

Loss Function

The loss function minimizes the discrepancy between actual and predicted ratings:

$$L = \sum_{(u,i)\in k}\left(R_{u,i} - p_u^T q_i\right) + \lambda(\|p_u\|^2 + \|q_i\|^2), \quad (2)$$

where k is the set of user-item pairs with ratings, λ is the regularization parameter, which prevents overfitting, p_u is the latent factor vector for user u, q_i is the latent factor vector for item i, $R_{u,i}$ is the actual rating given by user u to item i, and $p_u^T q_i$ is the predicted rating.

The first term represents the sum of squared prediction errors, which measures the discrepancy between actual and predicted ratings. The second term is a regularization component that constrains the magnitude of the latent factor vectors, helping to prevent overfitting by discouraging excessively large parameter values.

Training Process–Stochastic Gradient Descent

1. Randomly initialize the latent factor vectors p_u and q_i for all users and items.
2. For each known rating (u, i):
   - Compute the predicted rating: $\hat{R}_{u,i} = p_u q_i$.
   - Calculate the error: $e_{u,i} = R_{u,i} - \hat{R}_{u,i}$, where $R_{u,i}$ is the actual rating and $\hat{R}_{u,i}$ is the predicted rating.
3. Update rules for the latent factor vectors:
   - For user u, factor f: $p_{u,f} \leftarrow p_{u,f} + \gamma \times \left(e_{u,i} \times q_{i,f} - \lambda p_{u,f}\right)$.
   - For item i, factor f: $q_{i,f} \leftarrow +\gamma \times \left(e_{u,i} \times p_{u,f} - \lambda q_{i,f}\right)$.

Where γ is the learning rate, and λ is the regularization parameter.

4. Repeat the process for multiple iterations until the loss function converges or a preset number of iterations is reached.

In the business field, recommender systems are widely used to enhance user experience. For example, Netflix uses a variety of machine learning models to process data and generate recommendations, such as clustering and dimensionality reduction algorithms, linear regression, logistic regression, MF, and hybrid models. Netflix emphasizes personalized experiences, starting the recommendation system as soon as a user logs in. To meet the user's needs, the system uses data from playback history, ratings, queues, searches, etc., to generate relevant recommendations, while also helping users discover content they are interested in within a vast online database (Amatriain & Basilico, 2015).

MoocRec is a website that recommends courses to users to help them acquire the skills needed for their ideal jobs. Its recommendation engine combines MF models with CF algorithms and utilizes information from external resources to make the most suitable recommendations for users. First, the system gathers information from users, such as their learning background, ideal jobs, and required skills, to understand their needs and goals. Next, MoocRec integrates a search engine that uses web content mining techniques to automatically retrieve MOOC courses from providers like edX and Coursera, enabling the system to access a vast amount of course information. The MoocRec database then builds relationships between jobs and skills, as well as skills and courses, allowing the system to recommend courses that help users acquire the skills required for their ideal jobs (Symeonidis & Malakoudis, 2016).

In the education field, IT108 was developed in response to the implementation of the 108 curriculum guidelines. Its goal is to create an automated web-based online learning platform for the information technology field. It uses text mining technology to analyze curriculum documents, automatically extracting and analyzing learning content and descriptions, which are then stored in a hierarchical structure in a database. The platform also uses the YouTube Data API to retrieve a large number of instructional videos related to these keywords. It ranks the videos according to the

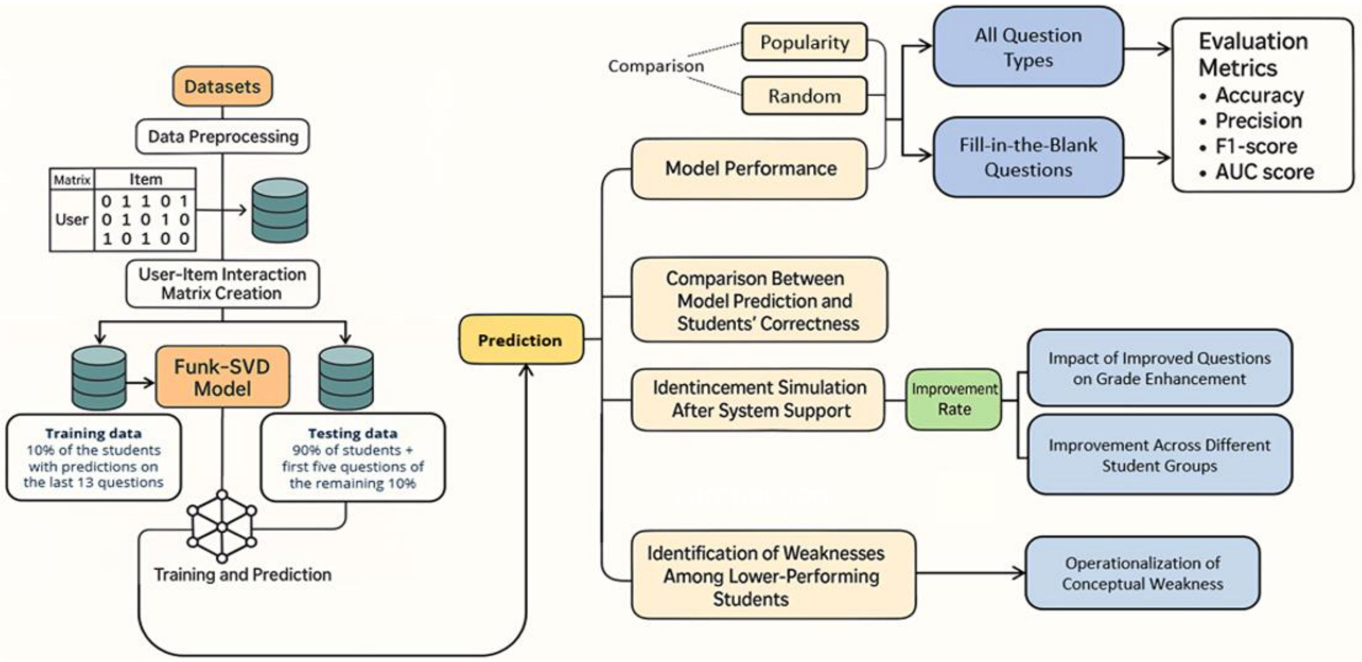

Figure 1. Methodology framework (Source: Authors' own elaboration)

corresponding curriculum content and keyword weight for each unit, storing the most relevant videos in the database. Furthermore, IT108 records each user's learning progress, such as notes, video viewing, and added items, enabling users to track their learning history and motivating independent learning (Lu, 2022). Overall, IT108 creates automated learning content based on the 108 curriculums, integrates YouTube's rich learning resources, and establishes an online learning environment. Through interactive features and a personalized recommendation system, it helps students learn information technology knowledge more effectively and assists teachers in lesson planning.

In a nutshell, recommendation systems based on Funk-SVD are currently widely used in the business field. While they have been introduced in education, the field is still in the early stages of development. In higher education, calculus courses form the foundation of many disciplines. Learning calculus enhances students' abstract thinking, logical reasoning, and mathematical modeling abilities. After understanding the abstract concepts of calculus, students also need good computational skills. Mathematics teachers often rely on experience in teaching. Via using recommender systems to analyze students' learning patterns, teachers can systematically understand students' learning states and design the most suitable teaching content for them.

Building on the contributions of these systems, the present study extends the application of recommendation technologies to the field of university-level calculus education. While previous research such as MoocRec and IT108 has focused on course retrieval and curriculum-based content recommendations, this study uniquely leverages Funk-SVD to analyze actual student performance data at the item level. By identifying conceptual weaknesses and predicting students' future performance, the system moves beyond content matching toward a more diagnostic and intervention-oriented approach. This not only bridges the gap between data mining and classroom instruction but also responds to the increasing need for personalized learning support in foundational STEM courses such as calculus.

METHODOLOGY

Our primary objective is to develop and evaluate a predictive recommendation system that supports individualized calculus learning by identifying students' conceptual weaknesses and simulating targeted intervention outcomes. This methodological framework is guided by three core objectives:

1. To construct a reliable user-item interaction matrix using student response data
2. To train and validate the Funk-SVD model for predicting individual performance on calculus problems
3. To evaluate model effectiveness through comparisons with baseline approaches and simulations of post-intervention improvement

As illustrated in **Figure 1**, the methodology begins with data preprocessing and matrix construction, followed by model training and prediction. Predictions are then analyzed through four evaluation paths: model accuracy (**RQ1**), impact on simulated learning gains (**RQ2**), benefit across student groups (**RQ3**), and identification of topic-level weaknesses (**RQ4**). These evaluation paths collectively frame the analytical structure of the study.

2. Let f, g be functions for which the range of g is in the domain of f (so that $f \circ g$ makes sense), we always have for any c in the domain of g that $\lim_{x \to c} f(g(x)) = f(\lim_{x \to c} g(x))$ provided $\lim_{x \to c} g(x)$ exists.

Ans. False.

5. Suppose f is a function of the two variables x, y and in term, each of the x and y is a function of both u and v, then f is a function of u and v. Given that $\frac{\partial f}{\partial u}(a,b) = 9$, $\frac{\partial f}{\partial x}(x(a,b), y(a,b)) = 2$, $\frac{\partial x}{\partial u}(a,b) = 4$, $\frac{\partial y}{\partial u}(a,b) = 9$. Find $\frac{\partial f}{\partial y}(x(a,b), y(a,b))$.

Answer: $\frac{1}{9}$

Figure 2. Examples of true/false questions and fill-in-the-blank questions (Source: Calculus examination items administered by the case university)

Data Sets

The data used in this research were obtained from a de-identified secondary database collected at a case study university in Taiwan, comprising student responses from calculus courses conducted over two consecutive semesters during the academic years 2022 and 2023. The dataset included the performance records of over 850 first-year undergraduate students enrolled in compulsory calculus courses across various majors in science, engineering, and technology-related disciplines.

Since the data were anonymized prior to analysis and no personal identifiable information was accessible, the study involved the analysis of pre-existing, de-identified data without any direct interaction with participants. As such, this research does not constitute human subjects research under the common rule (45 CFR 46.102) and falls within the domain of educational data mining (EDM), focusing on identifying performance patterns and diagnosing conceptual weaknesses to provide targeted practice recommendations.

In EDM, sample size is often regarded as a key determinant of machine learning model performance. Addressing the feasibility of small-sample modeling, Zohair and Mahmoud (2019) conducted an empirical study demonstrating that a predictive model with statistically significant accuracy can be effectively developed using only 38 student records. Through visualization and clustering techniques, critical predictors were identified, enabling the construction of a reliable student performance classification model. The study employed leave-one-out cross validation, a method widely recognized for maximizing data utility and ensuring model robustness in small-sample contexts (Rao et al., 2008).

In contrast, the present study utilizes a substantially larger dataset comprising 850 student records, offering greater data representativeness and improved model stability. Building upon prior findings, this study affirms the feasibility of applying machine learning to educational outcome prediction and further strengthens its methodological rigor through the use of a sufficiently large and diverse sample.

The data set consists of responses to a total of 138 distinct calculus problems, systematically distributed across six quizzes conducted throughout the two semesters. Each quiz costs 150 minutes and comprises

- 10 true/false questions (**Figure 2**)
- 8 fill-in-the-blank questions (**Figure 2**)
- 5 essay-type questions

Each student's response to each problem was recorded as a binary value: correct (1) or incorrect (0). Problems were specifically designed by experienced calculus instructors to assess students' comprehension across various calculus topics such as differentiation, integration, critical points, curvature, multiple integral and integral volume.

Data Preprocessing

Dataset cleaning

Data preprocessing began with the removal of attributes that were irrelevant to our predictive modeling objectives, such as instructor names, academic years, and degrees, to ensure clarity and consistency in the dataset. Moreover, to improve the precision and interpretability of predictions, essay-type questions were excluded from the analysis due to complexities associated with partial scoring. Consequently, the experiments utilized only the 10 true/false questions and 8 fill-in-the-blank questions per quiz. The essential attributes retained for the training and evaluation of the predictive model are summarized in **Table 1**.

Thus, the dataset consists of approximately 850 students × 108 questions (18 per quiz × 6 quizzes), yielding over 91,800 student-problem interaction records. These records serve as the input for the CF algorithm, enabling the model to learn latent factors that

Table 1. Main attributes of the data set

Attributes	Data type	Details
Student names	Nominal	Student identifiers (e.g., student1, student2, ...)
Student ID	Nominal	Unique identifiers assigned to students (e.g., ID001, ID002, ...)
Quiz number	Ordinal	Quiz sequences (1-6)
Question number	Ordinal	Individual problem numbers (1-23 per quiz)
Answer correctness	Nominal	Performance indicator: 1 (correct), 0 (incorrect)

student_id	question_1	question_2	question_3	question_4	question_5
id_1	0.0	0.0	0.0	1.0	0.0
id_2	0.0	1.0	0.0	1.0	0.0
id_3	0.0	0.0	1.0	1.0	1.0
id_4	0.0	1.0	1.0	0.0	1.0
id_5	0.0	0.0	0.0	1.0	0.0
id_6	0.0	0.0	0.0	1.0	0.0

Figure 3. Part of user-item interaction matrix (Source: Authors' own elaboration)

represent both student proficiency and question difficulty.

User-item interaction matrix creation

To effectively apply CF methods, the original dataset was transformed into a user-item interaction matrix, as exemplified in **Figure 3**. In this matrix structure, each student is represented as a user, each calculus problem is treated as an item, and the binary correctness values (correct or incorrect) indicate interactions between users and items. This format is suitable for implementing the Funk-SVD model, as it allows for the efficient extraction of latent patterns that reveal students' problem-solving capabilities and conceptual weaknesses in calculus.

Training and testing data split

The processed data was subsequently divided into distinct training and testing datasets to assess the predictive capability of our model rigorously:

- **Training set:** Includes data from 90% of students (approximately 765 students) and the first five questions answered by the remaining 10% (approximately 85 students).
- **Testing set:** Includes the last 13 questions from each quiz, as answered by the remaining 10% of students (approximately 85 students).

The inclusion of the first five questions in the training data for students in the testing subset was strategically determined through preliminary experimental analysis. These initial problems were found to provide sufficient data points for accurate prediction without introducing significant bias.

Model: Funk-SVD

Funk-SVD, a variant of SVD tailored specifically for recommender systems, decomposes the user-item interaction matrix into lower-dimensional latent factor representations for both users and items. This process captures the underlying structures in the interaction data, enabling the model to estimate a "preference" score for unseen user-item pairs. In this study, the "preference" score represents the likelihood that a student will answer a given problem correctly, serving as a proxy for predicted performance. The primary formula employed by Funk-SVD is shown in Eq (3):

$$s(u,i) = b_{ui} + \sum_{f=1}^{k} U_{uf} I_{if}, \quad (3)$$

where $s(u,i)$ is the predicted likelihood score for user u on item i, b_{ui} is baseline prediction (global average combined with user and item biases), U_{uf} is latent factor vector of user u, I_{if} is latent factor vector of item i, and k is number of latent factors (hyperparameter).

In our context, this prediction score corresponds to the estimated probability that a student will answer a specific calculus problem correctly, serving as the foundation for our targeted recommendation system.

Tools: The "Surprise" Library

In this study, we implemented the Funk-SVD model using Python, leveraging Google Collaboratory as a cloud-based environment to ensure both computational efficiency and reproducibility. To facilitate the development and evaluation of our recommendation system, we employed the surprise library, an open-source Python toolkit specifically designed for building and analyzing recommender systems based on CF techniques. Surprise provides a streamlined and modular interface that supports a wide range of algorithms, including Funk-SVD, and offers built-in functionalities for data loading, training/testing splitting, cross-validation, and performance evaluation. Its native support for parameter tuning and evaluation metrics such as RMSE and MAE made it particularly well-suited for our experimental framework. By integrating Surprise into our implementation, we were able to efficiently train and test the Funk-SVD model on our dataset, perform hyperparameter optimization, and evaluate predictive performance in a reproducible and scalable manner.

Training and Prediction

We trained the Funk-SVD model using an MF framework implemented via the surprise library in Python. The training dataset consisted of student-problem interaction records, where a score of 1 indicated

Table 2. Accuracy of different thresholds

Threshold	Accuracy
0.40	0.728
0.45	0.731
0.50	0.734
0.55	0.726
0.60	0.717

a correct response and 0 indicated an incorrect response. Model hyperparameters were set as follows: the number of latent factors was 50, the number of training epochs was 40, and the learning rate was set to 0.02. These parameters were optimized through cross-validation to balance prediction accuracy and model generalizability.

Upon completion of training, the model output was a prediction matrix that estimated the probability of correctness for each student-problem pair in the test dataset. Predicted scores ranged from 0 (predicted incorrect) to 1 (predicted correct).

To enable binary classification, we applied a fixed decision threshold. If prediction is larger than threshold, it would be classified as "correct"; Otherwise, it would be classified as "incorrect".

- Predictions ≥ threshold were classified as "correct"
- Predictions < threshold were classified as "incorrect"

To determine the optimal threshold for the model's predictions, multiple thresholds were systematically evaluated based on their predictive accuracy. The experimental results (**Table 2**) indicate that a threshold of 0.5 yielded the highest accuracy of 73.1%. This threshold was consequently selected for subsequent analyses to maximize the model's predictive performance.

Evaluation Metrics

To comprehensively evaluate the model's predictive performance, we utilized four established metrics commonly applied in classification problems:

- **Accuracy:** The proportion of correctly predicted outcomes out of the total number of predictions. The value is between 0.000 and 1.000, the closer to 1 the better.
- **Precision:** The proportion of true positive predictions out of all positive predictions made by the model. The value is between 0.000 and 1.000, the closer to 1 the better.
- **F1-score:** The harmonic means of precision and recall, providing a balanced assessment of the model's performance. The value is between 0.000 and 1.000, the closer to 1 the better.
- **Area under the curve (AUC) score:** A measure derived from the receiver operating characteristic (ROC) curve that evaluates the model's ability to distinguish between positive and negative classes. The value is between 0.000 and 1.000, the closer to 1 the better.

The combination of these metrics provides an extensive evaluation, allowing for robust assessment of the model's suitability and predictive capability in real-world educational settings (Han et al., 2011; Romero & Ventura, 2010).

Operationalization of Conceptual Weakness

In this study, a student's conceptual weakness is operationally defined based on the set of recommended items generated by the Funk-SVD model–specifically, the *N* items with the lowest predicted success probabilities. Each item was pre-tagged by domain experts with one or more underlying concepts, covering both core calculus topics (e.g., limits, derivatives, integrals) and prerequisite knowledge areas (e.g., algebraic manipulation, trigonometric identities, and coordinate geometry). The inter-rater agreement on these concept tags reached 97%, indicating a high level of internal consistency.

By aggregating the concept tags associated with each student's top-*N* recommended items, we inferred personalized areas of conceptual difficulty. A student's low predicted success on these items reflects not only deficiencies in isolated content knowledge but also challenges in applying mathematical concepts within structured problem-solving contexts. This integrated operationalization links students' weaknesses not only to specific topical gaps but also to broader cognitive competencies rooted in prerequisite knowledge. Such an approach aligns with prior research highlighting the foundational importance of algebra, trigonometry, and analytic geometry in the learning of calculus.

Importantly, our definition of "weakness" is data-driven–grounded in individual prediction patterns–rather than based solely on error counts or instructor judgment. It captures both localized struggles with particular topics and more pervasive difficulties in constructing the conceptual foundations essential for advanced mathematical reasoning.

RESULTS

Model Performance

To evaluate the performance of our Funk-SVD model, we compared its results with two standard baseline approaches: random prediction and popularity prediction.

- **Random prediction:** This approach predicts correctness randomly for each question, serving as a reference point for evaluating the model's effectiveness.

Table 3. Performance of each approach (all question types)

	Accuracy	Precision	F1-score	AUC score
Funk-SVD	0.734	0.777	0.794	0.701
Popularity	0.712	0.727	0.795	0.649
Random	0.502	0.641	0.560	0.504

- **Popularity prediction:** This method predicts correctness based on the correct answer rate of each question in the training data. Specifically, if a question's correct answer rate exceeds 0.5, the model predicts that all students will answer it correctly; otherwise, it predicts all responses as incorrect. This approach is commonly used by instructors to identify commonly misunderstood topics and guide instructional focus.

Performance on all question types

As shown in **Table 3**, the Funk-SVD model consistently outperformed both baseline approaches across all evaluation metrics. Specifically, the Funk-SVD model achieved an accuracy of 73.4%, precision of 77.7%, F1-score of 0.794, and AUC of 0.701, demonstrating strong predictive accuracy.

In comparison, the random baseline performed near chance level, as anticipated, whereas the popularity-based approach showed moderate yet notably inferior performance compared to Funk-SVD. The AUC score further underscores the model's superior ability to distinguish between correct and incorrect responses. Collectively, these findings support the robustness of the Funk-SVD model and highlight its suitability for practical implementation in personalized educational systems. Future work may explore statistical significance testing and evaluate prediction consistency across different student subgroups. **Figure 4** presents the ROC curves for the three models: the Funk-SVD model, popularity prediction, and random prediction. The ROC curve graphically represents each model's ability to differentiate correctly predicted student responses from incorrect ones by plotting the true positive rate against the false positive rate. Notably, the Funk-SVD model (represented by the blue line) attained the highest AUC (AUC = 0.701), indicating superior predictive accuracy. The popularity prediction model followed with a moderate AUC of 0.649, whereas the random prediction performed near chance level with an AUC of 0.504, equivalent to a no-skill classifier. These results further emphasize the robust predictive capability of the Funk-SVD model in accurately classifying student responses.

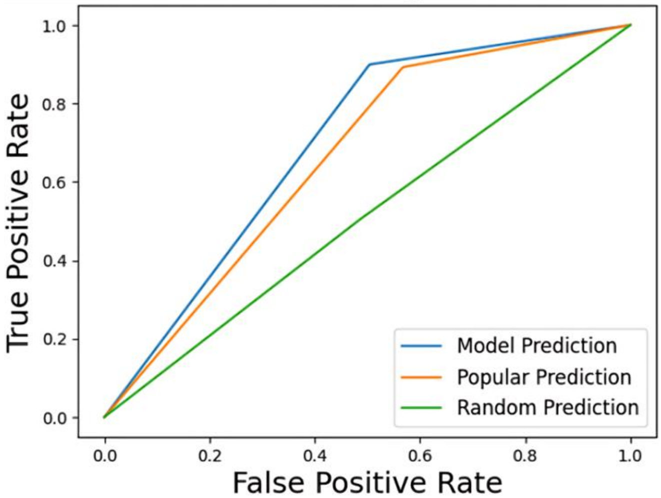

Figure 4. ROC curve comparison (all question types) (Source: Authors' own elaboration)

Table 4. Performance of each approach (fill-in-the-blank questions)

	Accuracy	Precision	F1-score	AUC score
Funk-SVD	0.694	0.680	0.690	0.694
Popularity	0.666	0.638	0.678	0.667
Random	0.491	0.476	0.478	0.491

Performance on fill-in-the-blank questions

To reduce the influence of random guessing, particularly prevalent in multiple-choice or true/false formats, we conducted a focused evaluation using fill-in-the-blank questions. These question types typically require students to recall and construct responses based on their understanding, making them a more valid indicator of actual mastery.

For this analysis, we selected 10 fill-in-the-blank questions per student as dataset. The dataset was split into training and testing sets, with the final three questions from each quiz reserved for testing. This setup was designed to simulate realistic learning trajectories while preserving temporal order.

Compared to the full dataset evaluation (**Table 3**), the slightly lower accuracy and AUC in **Table 4** reflect the increased difficulty and cognitive demand of fill-in-the-blank questions. Despite this, the Funk-SVD model remained the top performer across all metrics, suggesting that its predictive strength is not limited to patterns in high-frequency responses or superficial correctness. Rather, it demonstrates adaptability in identifying conceptual gaps that emerge in more open-ended or recall-based formats.

The Funk-SVD model maintained its superior predictive performance even in scenarios specifically designed to minimize the impact of guessing, as indicated in **Table 4**. In this targeted evaluation using fill-in-the-blank questions, the model achieved an accuracy of 69.4% and an AUC of 0.694. These results further validate the robustness of the Funk-SVD model, underscoring its capability to detect genuine conceptual misunderstandings rather than superficial response patterns. Such diagnostic accuracy supports the model's potential as a valuable tool for personalized learning interventions, particularly in contexts demanding higher-order cognitive processing.

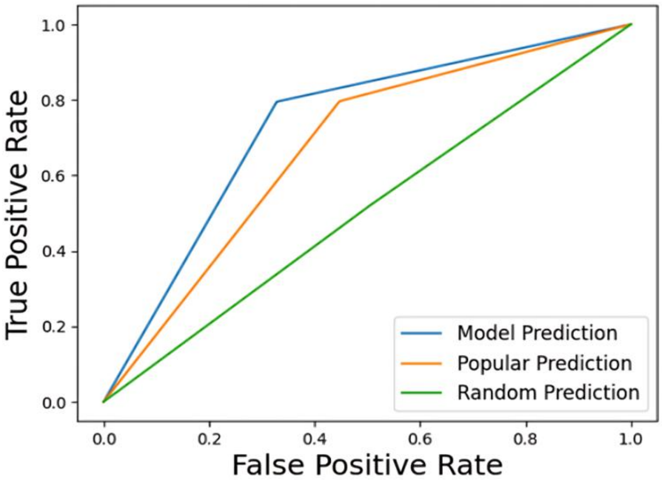

Figure 5. ROC curve comparison (fill-in-the-blank questions) (Source: Authors' own elaboration)

The evaluation of fill-in-the-blank questions further supports the robustness and effectiveness of the Funk-SVD model, especially considering that these question types are inherently more challenging and less susceptible to correct answers through guessing, as demonstrated in **Figure 5**. The ROC curve analysis highlights that the Funk-SVD model consistently achieves superior predictive performance (AUC = 0.694) compared to both the popularity prediction (AUC = 0.667) and random prediction (AUC = 0.491) models. This superior performance indicates the model's strength in distinguishing between genuine conceptual understanding and mere guessing behavior. Consequently, the Funk-SVD model shows significant potential for application in formative assessments, enabling educators to accurately identify and address students' conceptual weaknesses through targeted instructional interventions.

Comparison Between Model Prediction and Students' Correctness

To further evaluate the model's performance, we compared its prediction accuracy with the correct rate of each question. The correct rate represents the proportion of students who answered a particular question correctly, while the prediction accuracy measures how closely the model's predictions align with actual student outcomes for each question.

The comparison between predicted accuracy and actual student correct rates across all questions indicates that, in most cases, the Funk-SVD model's predictions closely align with or surpass the observed correct rates, as illustrated in **Figure 6**. This alignment underscores the model's ability to accurately identify patterns within student responses and reliably predict performance outcomes. Particularly noteworthy is the model's enhanced predictive accuracy for questions exhibiting very high or very low correct rates, highlighting its strength in effectively capturing areas where students commonly excel or struggle. These findings reinforce the potential effectiveness of the Funk-SVD model in targeted academic interventions.

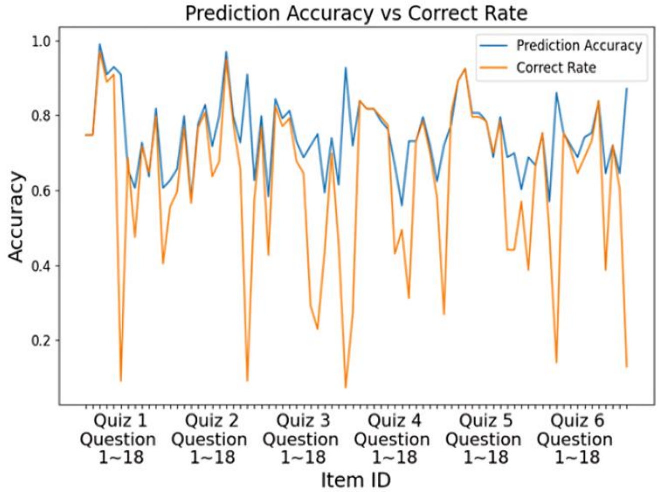

Figure 6. Comparison between prediction accuracy and correct rate (x-axis represents the 18 questions of each quiz & y-axis represents the prediction accuracy and correct rate) (Source: Authors' own elaboration)

However, the model exhibits lower performance for questions with a correct rate between 0.4 and 0.6. This observation suggests that the model's predictions are sensitive to question difficulty, particularly for moderately challenging questions. For items in this range, the model's accuracy dropped to approximately 0.5~0.6, indicating a decline in predictive reliability. Prior research in educational assessment has shown that questions with moderate difficulty levels often exhibit greater variability in student responses, potentially due to a mixture of partial understanding, guessing, or concept transfer difficulties (Mislevy et al., 2017; Wang & Chen, 2019).

Improvement Simulation After System Support

Figure 4 and **Figure 5** provide evidence that the Funk-SVD model effectively identifies individual students' weakest areas in calculus. The primary goal of our approach is not to modify the assessment items, but rather to use predictive modeling to pinpoint each student's conceptual gaps and recommend targeted practice accordingly. By aligning remediation with the system's predictions, we aim to foster meaningful learning and improve performance on similar problem types.

To evaluate the potential impact of these recommendations, we conducted a simulation modeling post-intervention outcomes. We assumed that students, upon receiving personalized recommendations, would review the predicted weak concepts and subsequently answer those questions correctly. In the simulation, these targeted items were marked as correct (answer = 1) to reflect the expected benefit of personalized intervention.

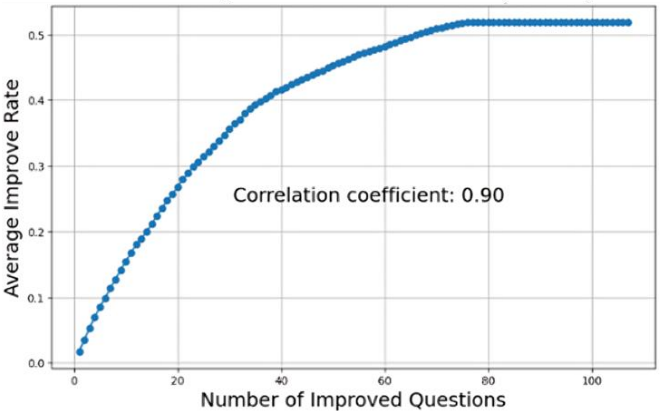

Figure 7. Students' grade improvement vs. number of improved questions (Source: Authors' own elaboration)

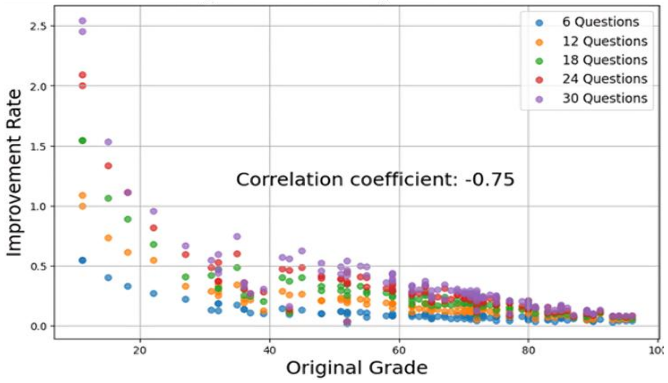

Figure 8. Original grade vs. improve rate (x-axis represents the original grades that the students got on the quiz & y-axis represents the improvement rate after using the system) (Source: Authors' own elaboration)

Improvement rate

The improvement rate in this study quantifies the relative gain in student performance resulting from a simulated personalized intervention based on the Funk-SVD model. Specifically, it measures the percentage increase in the number of correct responses after targeted practice on the weakest predicted concepts, relative to the original score. This metric serves as an indicator of the potential effectiveness of model-driven remediation and reflects how much personalized guidance–enabled by the Funk-SVD model–could enhance a student's mastery if applied in practice. In this study, an improvement rate exceeding 25% is regarded as indicative of meaningful learning gains (Koedinger et al., 2015; Slavin, 2002). The improvement rate is defined as follows:

$$Improvement\ rate = \frac{Y_i - X_i}{X_i}, \quad (4)$$

where X_i **(original grade)** is oiginal number of correct answers before intervention (maximum = 108) and Y_i **(new grade)** is number of correct answers after simulated improvement.

Impact of improved questions on grade enhancement

The strong linear correlation between the number of improved questions and the overall improvement rate provides further validation of the system's effectiveness in enhancing student performance. As illustrated in **Figure 7**, there is a pronounced diminishing return effect, where interventions at earlier stages produce greater relative improvements. Specifically, the marginal benefit per additional question corrected is notably higher when fewer questions have been addressed; For example, the slope of improvement at 30 corrected questions is considerably steeper than at 70 questions, indicating that targeted interventions are most efficient during the initial phase of practice. This trend is further substantiated by a high correlation coefficient (r = 0.90), demonstrating both the strength and consistency of the relationship across students and confirming the robustness of the observed pattern. These findings reinforce the practical value of early targeted interventions in personalized learning strategies.

The results also reveal that the improvement rate increases with the number of corrected questions but begins to plateau around 80 questions. This suggests diminishing returns and highlights the system's ability to prioritize the most critical items for improvement. Consequently, substantial grade enhancements can be achieved with minimal intervention, supporting efficient and personalized learning strategies.

Improvement across different student groups

While a moderate negative correlation was observed between original grades and improvement rates (r = -0.75), further analysis revealed nuanced patterns across different student groups. As illustrated in **Figure 8**, each data point represents an individual student's improvement following five targeted interventions with varying numbers of recommended questions (6, 12, 18, 24, or 30).

Firstly, students with lower original scores (below 40) demonstrated the most significant improvement rates, highlighting the particular effectiveness of targeted interventions for learners struggling with fundamental concepts. For instance, these students frequently achieved improvement rates exceeding 50% when 30 questions were assigned, and rates surpassing 25% even when only 6 items were provided.

Moreover, within this lower-performing subgroup, a high degree of variability in improvement rates was observed. Specifically, the model implemented five distinct intervention intensities, and students who received larger sets of recommended questions tended to show greater simulated improvement.

Secondly, higher-performing students exhibited diminishing returns from similar interventions, reflecting their fewer conceptual weaknesses requiring remediation.

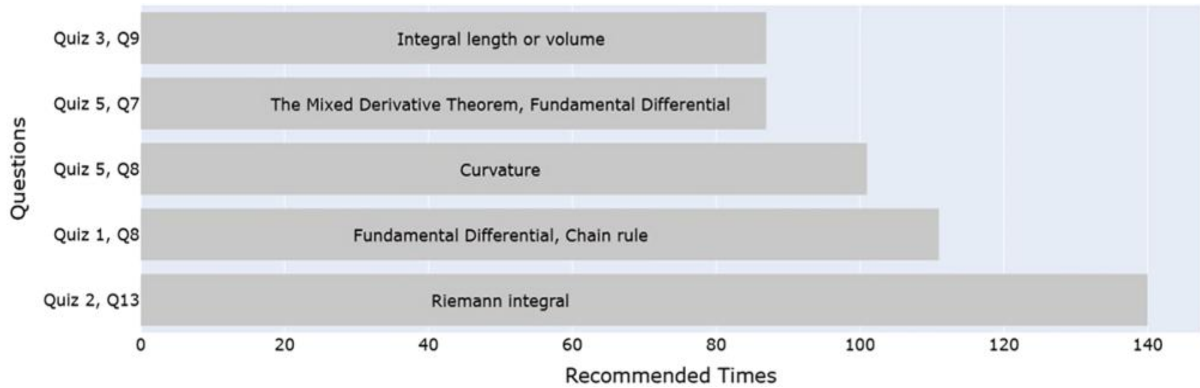

Figure 9. Top-5 recommended topics for lower-performing students (Source: Authors' own elaboration)

Collectively, these findings emphasize the system's capability to deliver substantial benefits to students most in need while also retaining applicability to higher-performing learners. Thus, the results underscore the practical value of the Funk-SVD model in facilitating differentiated instruction, enabling educators to efficiently provide targeted and impactful interventions across a diverse range of student abilities.

Identification of Weaknesses Among Lower-Performing Students

To generate actionable instructional predictions, we used the Funk-SVD model to identify the N = 10 problems with the lowest predicted success probabilities for each student. These items reflect the most probable areas of conceptual weakness and were pre-tagged by domain experts with their corresponding calculus concepts. This enabled concept-level diagnostic inference and allowed the system to generate personalized recommendation sets aligned with each student's most urgent learning needs.

Given limitations on instructional time and teaching resources, we conducted a focused analysis on the lowest-performing 30% of students ($n \approx 255$). For each student, the ten lowest-ranked problems were extracted and analyzed. **Figure 9** presents the five most frequently recommended problems and their associated conceptual tags across this subgroup.

The statistical analysis of these aggregated recommendations revealed a multi-level structure of conceptual difficulty among lower-performing students (**Figure 9**): At the initial learning stage, a substantial proportion of students had difficulty mastering foundational calculus topics such as Riemann integral, fundamental integral, fundamental differential, and chain rule. These persistent difficulties suggest a lack of success in establishing basic skills in differentiation and integration. As students move into the intermediate stage, additional challenges emerged in tasks requiring geometric reasoning and coordinate transformations. Problems involving arc length, volume of revolution, and curvature appeared frequently in the recommendation sets, indicating that many students lacked the spatial intuition and flexibility needed to solve geometry-intensive calculus problems. At the advanced stage, students struggled with items involving the mixed derivative theorem, often due to an insufficient understanding of the continuity conditions required for interchanging the order of partial differentiation–as formalized in Schwarz's theorem. Such misunderstandings pose significant barriers to engaging with multivariable calculus rigorously.

Notably, many of the most frequently recommended items were also associated with prerequisite knowledge domains, such as algebraic manipulation, function transformation, and coordinate geometry. This observation is consistent with findings from related work on the teaching and learning of calculus, which emphasize that deficiencies in foundational mathematics–rather than purely calculus-specific misunderstandings–often account for students' persistent struggles. Such gaps in fluency fundamentally hinder students' ability to engage with abstract and integrative problem-solving tasks in calculus.

Taken together, these findings highlight the diagnostic value of the Funk-SVD model–not only in identifying surface-level topic weaknesses, but also in uncovering deeper learning obstacles often overlooked by conventional assessments. The model's predictive output offers a robust, data-informed foundation for strategic instructional planning, empowering educators to prioritize high-impact concepts, allocate limited resources more effectively, and provide targeted support to those students most in need.

CONCLUSION

This study presents a novel approach to identifying and addressing student weaknesses in calculus through a data-driven recommendation system. By leveraging the Funk-SVD model, a widely used CF technique, we successfully predicted student performance on individual calculus problems and provided targeted recommendations for improvement.

The Funk-SVD model demonstrated superior performance compared to baseline approaches, including random prediction and popularity prediction, across various metrics such as accuracy, precision, F1-score, and AUC. Furthermore, focused analyses on fill-in-the-blank questions highlighted the model's robustness in minimizing the impact of guessing and accurately identifying student weaknesses.

Through simulation experiments, we demonstrated the system's practical utility in supporting student learning. By recommending weak topics and simulating improvement on low-performing questions, the system showed significant potential for enhancing students' grades, especially for underperforming students. The analysis revealed that addressing fewer than 80 critical questions led to substantial improvements, with the greatest benefits observed among students with lower initial grades.

Additionally, the analysis of the most frequently recommended problems revealed consistent conceptual challenges among lower-performing students–particularly in topics such as volume integration, curvature, and Riemann sums. These results provide instructors with clear directions on which topics to prioritize for targeted remediation, enabling a more focused and efficient allocation of instructional time and resources.

These findings also carry broader implications for educational planning. By improving foundational understanding early in students' academic journey, such data-informed interventions may not only improve course-level outcomes but also reduce failure rates and support long-term academic continuity. In particular, this approach may help address issues of academic disengagement or attrition among first-year students, thereby enhancing their progression into the second year of university and beyond.

For instructors, the system provides valuable insights into commonly misunderstood topics, enabling data-driven adjustments to teaching strategies. For students, the system delivers personalized learning pathways, empowering them to focus on areas requiring the most attention.

However, this study did not involve a follow-up teaching experiment to directly assess how students respond to and benefit from the recommended practice in real classroom settings. As such, the system's effects on actual learning outcomes remain to be empirically validated. Future implementations could involve embedding the system into learning management systems or instructor dashboards to facilitate real-time monitoring and instructional decision-making.

In future work, we aim to refine the system further by incorporating additional contextual data, such as problem difficulty levels, learning styles, and engagement metrics, to improve prediction accuracy and expand its adaptive capabilities. Additionally, applying this framework to other subject areas could provide a scalable solution for personalized learning across diverse academic disciplines. By bridging advanced recommendation algorithms with educational practice, this study demonstrates the potential of technology to transform traditional learning environments and foster academic success.

Author contributions: C-CH: conceptualization, methodology, data analysis, and manuscript writing & **C-YH:** conceptualization, literature review, theoretical framework, supervision, funding acquisition, and manuscript revision. Both authors agreed with the results and conclusions.

Funding: This study was supported by [Approach to the deconstruction and deepening of calculus knowledge: A case study of university freshmen across academic years, 113-2423-H-008-002-MY2] from National Science and Technology Council.

Acknowledgments: The authors would like to thank National Central University for providing access to the de-identified secondary data used in this study.

Ethical statement: The authors stated that the study utilized de-identified learning records from previous calculus courses at a university, classified as secondary data. The dataset contained no personally identifiable information, and was used exclusively for data analysis to examine the effectiveness and educational application of the recommendation system. No intervention or interaction with human subjects was involved.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Amatriain, X., & Basilico, J. (2015). Recommender systems in industry: A Netflix case study. In *Recommender systems handbook* (pp. 385-419). Springer. https://doi.org/10.1007/978-1-4899-7637-6_11

Ancheta, C. M. D. (2022). An error analysis of students' misconceptions and skill deficits in precalculus subjects. *Journal for Educators, Teachers and Trainers, 13*(5), 283-295. https://doi.org/10.47750/jett.2022.13.05.026

Angco, R. J. N. (2021). Error analysis in integral calculus: A modified Newman's approach. *American Research Journal of Humanities Social Science, 4*(09), 38-52.

Aydin, U., & Ubuz, B. (2014). Predicting undergraduate students' mathematical thinking about derivative concept: A multilevel analysis of personal and institutional factors. *Learning and Individual Differences, 32*, 80-92. https://doi.org/10.1016/j.lindif.2014.03.023

Baker, B., Cooley, L., & Trigueros, M. (2000). A calculus graphing schema. *Journal for Research in Mathematics Education, 31*(5), 557-578. https://doi.org/10.2307/749887

Berggren, J. L. (2016). Calculus: Mathematics. In *Encyclopedia Britannica* (pp. 2-15).

Bi, X., Yan, S., & Gao, Y. (2019). Personalized recommendation model based on mini-batch gradient descent. *Computer Science and Application, 9*(04), 695-702. https://doi.org/10.12677/CSA.2019.94079

Billsus, D., & Pazzani, M. J. (1998). Learning collaborative information filters. In J. Shavlik (Ed.), *Proceedings of the 15th International Conference on Machine Learning (ICML)* (pp. 46-54). Morgan Kaufmann Publishers, Madison, WI.

Bressoud, D., Ghedamsi, I., Martinez-Luaces, V., & Törner, G. (2016). *Teaching and learning of calculus*. Springer. https://doi.org/10.1007/978-3-319-32975-8_1

Cappetta, R. W., & Zollman, A. (2013). Agents of change in promoting reflective abstraction: A quasi-experimental, study on limits in college calculus. *Journal of Research in Mathematics Education, 2*(3), 343-357. https://doi.org/10.4471/redimat.2013.35

Chang, T. K. (2021). Math students' misunderstandings and obstacles in learning limits of functions. *Taiwan Journal of Mathematics Education, 8*(2), 43-76. https://doi.org/10.6278/tjme.202110_8(2).002

Cheng, B. E. (2007). An experimental study on using Excel to assist calculus learning [Unpublished master's thesis]. National Changhua University of Education.

Cornu, B. (1981). Apprentissage de la notion de limite: Modèles spontanés et modèles propres [Learning the concept of limit: Spontaneous models and eigen models]. In *Actes du cinquième colloque du groupe internationale PME* (pp. 322-326). https://doi.org/10.4000/books.puv.5743

Cornu, B. (1983). *Apprentissage de la notion de limite: Conceptions et obstacles* [Learning the concept of limits: Conceptions and obstacles] [PhD dissertation, University of Grenoble].

Cornu, B. (1991). Limits. In D. Tall (Ed.), *Advanced mathematical thinking* (pp. 153-166). Kluwer. https://doi.org/10.1007/0-306-47203-1_10

de Vera, J. V., Peteros, E. D., Peconcillo Jr, L. B., Alcantara, G. A., Villarin, E. R., & Fulgencio, M. D. (2022). Assessing differences on students' attributes in mathematics based on their learning sessions. *Open Journal of Social Sciences, 10*(3), 170-185. https://doi.org/10.4236/jss.2022.103012

Dilling, F., & Witzke, I. (2020). The use of 3D-printing technology in calculus education: Concept formation processes of the concept of derivative with printed graphs of functions. *Digital Experiences in Mathematics Education, 6*, 320-339. https://doi.org/10.1007/s40751-020-00062-8

Ferrer, F. P. (2016). Investigating students' learning difficulties in integral calculus. *PEOPLE: International Journal of Social Sciences, 2*(1), 310-324. https://doi.org/10.20319/pijss.2016.s21.310324

Ferrini-Mundy, J., & Gaudard, M. (1992). Secondary school calculus: Preparation or pitfall in the study of college calculus? *Journal for Research in Mathematics Education, 23*(1), 56-71. https://doi.org/10.5951/jresematheduc.23.1.0056

Funk, S. (2006). Netflix update: Try this at home. *Sifter*. http://sifter.org/~simon/journal/20061211.html

Hammoudi, M. M., & Grira, S. (2023). Improving students' motivation in calculus courses at institutions of higher education: Evidence from graph-based visualization of two models. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(1), Article em2209. https://doi.org/10.29333/ejmste/12771

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques*. Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-381479-1.00003-4

Hurdle, Z. B., & Mogilski, W. (2022). The impact of prerequisites for undergraduate calculus I performance. *International Electronic Journal of Mathematics Education, 17*(3), Article em0696. https://doi.org/10.29333/iejme/12146

Idris, N. (2001). *Pedagogi dalam pendidikan matematik* [Pedagogy in mathematics education]. Utusan Publications.

Keene, K. A., Hall, W., & Duca, A. (2014). Sequence limits in calculus: Using design research and building on intuition to support instruction. *ZDM Mathematics Education, 46*(4), 561-574. https://doi.org/10.1007/s11858-014-0597-8

Kidron, I. (2020). Calculus teaching and learning. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 87-94). Springer. https://doi.org/10.1007/978-3-030-15789-0_18

Koedinger, K. R., Kim, J., Jia, J. Z., McLaughlin, E. A. & Bier, N. L. (2015). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In *Proceedings of the 2nd ACM Conference on Learning@ Scale* (pp. 111-120). ACM. https://doi.org/10.1145/2724660.2724681

Koren, Y. (2010). Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Transactions on Knowledge Discovery from Data, 4*(1), 1-24. https://doi.org/10.1145/1644873.1644874

Kunwar, R. (2021). A Study on low performing students perception towards mathematics: A case of secondary level community school students of Nepal. *Researcher: A Research Journal of Culture and Society, 5*(1), 125-137. https://doi.org/10.3126/researcher.v5i1.41384

Larsen, S., Marrongelle, K., Bressoud, D. M., & Graham, K. (2017). Understanding the concepts of calculus: Frameworks and roadmaps emerging from educational research. In J. Cai (Ed.), *Compendium for* *research in mathematics education* (pp. 526-550). National Council of Teachers of Mathematics.

Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing, 7*(1), 76-80. https://doi.org/10.1109/MIC.2003.1167344

López-Leyton, C., Aldana-Bermúdez, E., & Flórez-Laiseca, A. M. (2024). A socio-epistemological approach articulated with problem-solving in higher education: Teaching of integral calculus. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(12), Article em2548. https://doi.org/10.29333/ejmste/15704

Lu, W. H. (2022). *Automatic construction of web-based learning platform for IT domain of national education curriculum guide* [MA thesis, National Chi Nan University]. https://doi.org/10.6837/ncnu202200132

Mahadewsing, R., Getrouw, D., & Calor, S. M. (2024). Prior knowledge of a calculus course: The impact of prior knowledge on students' errors. *International Electronic Journal of Mathematics Education, 19*(3), Article em0786. https://doi.org/10.29333/iejme/14765

Martin, J. (2013). Difference between experts' and students' conceptual images of the mathematical structure of Taylor series convergence. *Educational Studies in Mathematics, 82*(2), 267-283. https://doi.org/10.1007/s10649-012-9425-7

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 3-62. https://doi.org/10.1207/S15366359MEA0101_02

Mkhatshwa, T. P. (2024). Best practices for teaching the concept of the derivative: Lessons from experienced calculus instructors. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(4), Article em2426. https://doi.org/10.29333/ejmste/14380

Parr, E. D. (2023). Undergraduate students' interpretations of expressions from calculus statements within the graphical register. *Mathematical Thinking and Learning, 25*(2), 177-207. https://doi.org/10.1080/10986065.2021.1943608

Pyzdrowski, L. J., Sun, Y., Curtis, R., Miller, D., Winn, G., & Hensel, R. A. (2013). Readiness and attitudes as indicators for success in college calculus. *International Journal of Science and Mathematics Education, 11*(3), 529-554. https://doi.org/10.1007/s10763-012-9352-1

Rao, R. B., Fung, G., & Rosales, R. (2008). On the dangers of cross-validation: An experimental evaluation. In *Proceedings of the 2008 SIAM International Conference on Data Mining* (pp. 588-592). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611972788.54

Rendle, S. (2010). Factorization machines. In *Proceedings of the 2010 IEEE International Conference on Data Mining* (pp. 995-1000). IEEE. https://doi.org/10.1109/ICDM.2010.127

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(6), 601-618. https://doi.org/10.1109/TSMCC.2010.2053532

Schraeder, M. R., Pyzdrowski, L. J., & Miller, D. A. (2019). The impact of prior exposure to calculus. *American Journal of Educational Research, 7*(3), 237-243. https://doi.org/10.12691/education-7-3-8

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*(7), 15-21. https://doi.org/10.3102/0013189X031007015

Sofronas, K. S., de Franco, T. C., Vinsonhaler, C., Gorgievski, N., Schroeder, L., & Hamelin, C. (2011). What does it mean for a student to understand the first-year calculus? Perspectives of 24 experts. *Journal of Mathematical Behavior, 30*(2), 131-148. https://doi.org/10.1016/j.jmathb.2011.02.001

Swinyard, C. (2011). Reinventing the formal definition of limit: The case of Amy and Mike. *The Journal of Mathematical Behavior, 30*(2), 93-114. https://doi.org/10.1016/j.jmathb.2011.01.001

Swinyard, C., & Larsen, S. (2012). Coming to understand the formal definition of limit: Insights gained from engaging students in reinvention. *Journal for Research in Mathematics Education, 43*(4), 465-493. https://doi.org/10.5951/jresematheduc.43.4.0465

Symeonidis, P., & Malakoudis, D. (2016). MoocRec. com: Massive open online courses recommender system. In *Proceedings of the Poster Track of the 10th ACM Conference on Recommender Systems*. ACM.

Thompson, P. W., & Harel, G. (2021). Ideas foundational to calculus learning and their links to students' difficulties. *ZDM–Mathematics Education, 53*(3), 507-519. https://doi.org/10.1007/s11858-021-01270-1

Ting, M. Y. (2018). Definite integral automatic analysis mechanism research and development using the "find the area by integration" Uunit as an example. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(7), 2883-2896. https://doi.org/10.12973/eurasia.2017.00724a

Wang, X., & Chen, Y. (2019). Modeling student response variability in adaptive learning environments. *Journal of Educational Measurement, 56*(2), 182-201. https://doi.org/10.1111/jedm.12202

Zhang, Y. (2022). An introduction to matrix factorization and factorization machines in recommendation system, and beyond. *arXiv*. https://doi.org/10.48550/arXiv.2203.11026

Zohair, A., & Mahmoud, L. (2019). Prediction of Student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education, 16*(1), 1-18. https://doi.org/10.1186/s41239-019-0160-3