Mentora ChatBot as an Intelligent Recommender System in Education

Olga Dukić[1[0009−0004−1694−9592]] and Marija Blagojević[1[0000−0003−4186−0448]]

University of Kragujevac, Faculty of Technical Sciences Čačak, Department of Information Technologies, Čačak, Serbia
email: dukicolga678@gmail.com, marija.blagojevic@ftn.kg.ac.rs

Abstract. The development and evaluation of the Mentora ChatBot, an educational conversational agent designed to provide personalized learner support, are presented. Mentora combines large language models (LLMs) with sequential modeling of user behavior, relying on a BiLSTM architecture alongside heuristic rules to generate recommendations aligned with learners' needs.
The system was developed and initially tested during preparatory classes for the final exam in mathematics and the Serbian language, conducted from February–June 2025 at a private educational institution. Twenty-two eighth-grade students participated, and more than 600 interaction logs were analyzed. Substantial gains in recommendation precision and acceptance were observed with the integration of the BiLSTM module, relative to a purely heuristic baseline.
An empirically grounded approach is introduced that couples LLM-driven dialogue with personalization of learning trajectories. Directions for further development are outlined, including the adaptation of recommendations to diverse learner profiles and the incorporation of explainable recommendations (explainable AI).

Keywords: Recommender systems · educational chatbot · BiLSTM · log analysis

1 Introduction

Contemporary education increasingly strives for a personalized approach to learning, wherein recommender systems and artificial intelligence (AI) play a growing role in monitoring progress and tailoring content to individual learner needs [1,5]. A particular place in this transformation is occupied by large language models (LLMs) and conversational agents, which enable dialog-based interaction, automated feedback, and intelligent guidance of the learning flow.

The development and evaluation of the **Mentora ChatBot**, an educational conversational assistant that uses a combination of LLMs, log analysis, and sequential modeling to generate learner-tailored recommendations, are presented. The system was developed as part of the broader *Mentora* educational platform, designed for deployment in real school settings and encompassing interactive lessons, quizzes, AI-based assistance, and learner progress tracking.

Research Paper
DOI: https://doi.org/10.46793/eLearning2025.171D
Part of ISBN: 978-86-89755-37-4

Mentora ChatBot combines explicit data (e.g., learners' answers to questions) with implicit signals from behavioral logs during learning sessions to produce personalized recommendations. Within the recommender module, a BiLSTM model is implemented to enable sequential analysis and tracking of learning trajectories over time. In addition, heuristic approaches are employed to obtain a hybrid effect by combining rule-based logic with data-driven learning. The current version of the BiLSTM model operates on synthetically generated data, whereas real user logs are in the process of being collected and expanded.

This study addresses the following research questions:

- How can large language models support the generation of educational recommendations within a conversational interface?
- To what extent does combining explicit data and behavioral logs contribute to higher-quality personalization?
- How can sequential models, such as a BiLSTM architecture, improve the dynamic adaptation of recommendations to the learner over time?

The remainder of this paper is organized as follows. Section 2 reviews relevant literature on educational recommender systems and conversational agents. Section 3 describes the architecture and functionalities of the Mentora system. Sections 4 and 5 detail the development of the recommender module and the analysis of log data, while Section 6 presents the experimental results and evaluation. Concluding remarks and directions for future development are provided in Section 7.

The Mentora system was developed during the 2024/2025 academic year within a private educational institution, as part of broader research into the application of AI in personalized learning. The objective is the enhancement of self-regulated learning by leveraging dialog-based interaction and recommendations grounded in prior learning experiences.

2 Theoretical Background and Literature Review

This section summarizes the theoretical foundations and current research directions that have shaped the development of the *Mentora ChatBot* system. The focus is on (i) personalization of learning through recommender systems, (ii) the role of conversational agents, (iii) analysis of usage logs and sequential modeling (KT/sequence-aware RS), and (iv) explainable recommendations and knowledge graphs. In this way, a conceptual framework is established for a hybrid approach (heuristics + BiLSTM within dialogue).

2.1 Recommender Systems in Education

Recommender systems (RS) in education are most commonly based on content-based filtering (CBF), collaborative filtering (CF), and hybrid approaches that combine multiple signals [9]. In CBF, profiles are formed from attributes of lessons/tasks, whereas CF relies on similarities derived from behavioral patterns across larger groups of learners. Hybrid approaches are particularly suitable in school settings due to data heterogeneity (different topics, tasks, learning styles) and the cold-start problem [5,1]. For working with web and usage logs, as well as evaluation metrics (*Precision*@k, *Recall*@k, NDCG@k), standard data-mining literature was followed [2,4,11]. In line with prior work advocating multi-layer architectures for educational recommender systems, the system was modularized into logging, evaluation, and recommendation components [10].

2.2 Conversational Recommender Systems (CRS)

Conversational recommenders (CRS) fuse recommendation generation with dialog: through multiple turns, preferences and constraints are elicited, explanations are provided, and suggestions are iteratively adapted. Surveys of CRS [3] and recent work in the area point to key components: a user model, a dialog management strategy, integration of background knowledge, and evaluation (offline and user studies). In education, CRS naturally align with *learning pathways*—dialog is used for diagnosis and the selection of next steps [12]. In the present approach, language understanding and generation are handled by an LLM, whereas recommendations are produced in a hybrid manner: heuristics for stability and a BiLSTM for sequential context.

2.3 Log Analysis and Sequential Modeling

Usage logs (interactions with lessons, tasks, and the bot) contain implicit patterns of progression, impasses, and knowledge transfer. Sequential modeling of these sequences (sessions, events, inter-event times) enables anticipation of the learner's next step. The seminal *Deep Knowledge Tracing* line of work has shown that LSTM models over task/answer sequences effectively predict performance [7]. Complementary surveys on sequence-aware recommender systems further support the use of sequential context for improved predictions [8]. In the same spirit, a BiLSTM is employed to couple the learner's "history" and "current context," with operation supported under limited real logs augmented by synthetic sequences (pilot phases and cold start). Empirical studies in learning analytics also document predictive value in real classroom logs [6].

2.4 Knowledge Graphs and Explainable Recommendations

Explainable recommendations are essential for trust and uptake, particularly in education, where both learners and teachers seek the "why" alongside the "what." Knowledge graphs (concepts, prerequisites, learning objectives) and a retrieval-augmented generation (RAG) layer over lessons provide a natural basis for explanations: "recommended because it is a prerequisite for X" or "because recent tasks in Y were challenging." Surveys of explainable RS [13] and work on graph-based learning paths indicate the effectiveness of these approaches in EdTech. In the current system, explanations are grounded in lessons and objectives (RAG), while explicit ontologies are planned for subsequent iterations.

3 System Description and Methodology

3.1 Architecture of the Mentora ChatBot System

Mentora ChatBot is a modular conversational system designed to enable a personalized educational experience in primary and secondary school contexts. The system was developed as part of the broader Mentora platform and is based on the integration of large language models (LLMs), an interaction-tracking layer, a heuristic recommendation layer, and a sequential model built on a BiLSTM architecture.

The architecture of the Mentora ChatBot system, which includes the LLM module, the lesson embedding space, the heuristic layer, answer evaluation, and the BiLSTM module for sequential modeling of learner behavior, is shown in Figure 1.

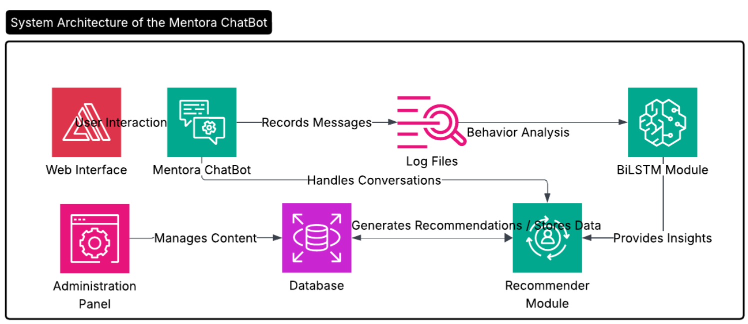

Fig. 1. Architecture of the Mentora ChatBot system

The system components include:

- **LLM module** — the OpenAI GPT-4o API is used for natural-language understanding and generation;
- **Lesson repository and embedding space** — lessons are represented by semantic vectors and tags and are indexed for relevant retrieval;
- **Answer evaluation system** — learner solutions are automatically analyzed (textual, visual, and audio input);
- **Recommendation system** — heuristics are applied based on learner behavior, with an integrated BiLSTM module proposing next learning steps.

A similar multi-layer architecture for e-learning hybrid recommenders has been described, and the present study focuses on the K–12 deployment setting [10].

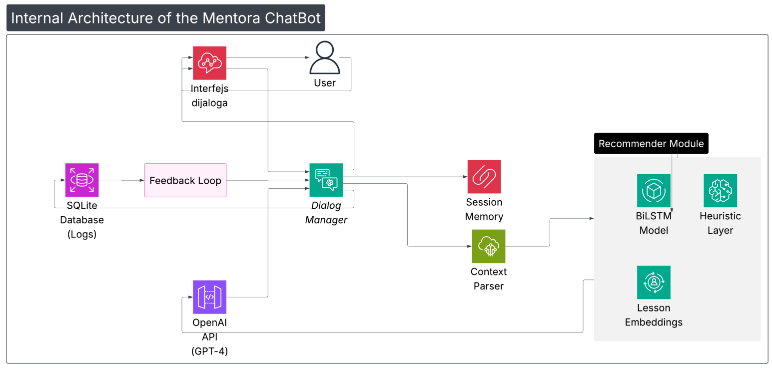

Fig. 2. Internal architecture of the Mentora ChatBot

3.2 Internal Architecture of the Mentora ChatBot

The internal architecture of the Mentora ChatBot is shown in Figure 2. The system consists of multiple interconnected modules organized around user interaction and personalized learning support. The key components are:

- **Input module** — textual, image, or voice input is received and forwarded to the system;
- **Context Parser** — each message is analyzed in the context of prior interactions, with intent and learning phase identified;
- **Dialog Manager** — the course of the conversation is managed and subsequent steps are determined;
- **Session Memory** — prior answers, recommendations, and bottlenecks are retained to ensure continuity;
- **Recommendation module** — individualized suggestions are generated using heuristics and a BiLSTM model;
- **Lesson Embeddings** — semantic retrieval of content is enabled;
- **LLM API layer** — natural-language responses are generated using the GPT-4o model;
- **Feedback loop** — behavioral data are leveraged to improve personalization.

This architecture enables Mentora ChatBot not only to function as a conversational agent but also as an adaptive tutor that learns about the user and monitors progress. The BiLSTM module contributes in particular to understanding behavioral patterns (e.g., recurring errors, returns to prior lessons), thereby producing recommendations aligned with the learner's pace and style.

3.3 Data Collection and Processing

Data are collected directly from learner interactions with the system and include:

- textual answers and questions;
- selected areas of interest;
- task-solving performance;
- session duration and activity order;
- visual inputs (images of tasks, drawings).

All data are stored in a relational database and, in parallel, converted into sequential series for training the BiLSTM module. In the initial phase, the model was tested on synthetic data that emulate typical learner behavior. Starting in September 2025, processing of real data is planned to improve performance and recommendation accuracy.

3.4 System Functionality

Mentora ChatBot integrates multiple functionalities for learning personalization and learner support:

- **Interactive tutor** — questions are answered, learners are guided through lessons, and assistance with problem solving is provided;
- **Recommendations** — subsequent lessons, tasks, and explanations are suggested;
- **Answer evaluation** — textual and visual responses are analyzed and feedback is provided;
- **Progress tracking** — activity is logged and learner progress is measured;
- **Adaptive feedback** — heuristics and the BiLSTM model are used to deliver intelligent recommendations.

The system was implemented using the following technologies:

- **Backend:** Django, PostgreSQL;
- **Frontend:** HTML, CSS, JavaScript, Bootstrap;
- **AI:** GPT-4o API, text-embedding-3-small, prompt engineering;
- **Recommender:** heuristics + BiLSTM in a TensorFlow/Keras environment.

Configuration and API keys are stored in an `.env` file, with scalability supported through a modular architecture.

3.5 Role of the Recommender Module in Dialogue

A modular structure is used to generate recommendations during dialog with the learner. When a need is expressed or a question is posed (e.g., "I do not know how to solve a fractions problem"), the context is recognized and the corresponding part of the recommender module is triggered.

Depending on the quantity and quality of prior interactions with the learner, the following is activated:

- a heuristic approach — in the case of a new user or insufficient data;

– the BiLSTM model — when sufficient information is available about the learner's prior behavior in the system.

The recommendation is then surfaced through the dialog, and multiple options are offered: the suggestion can be accepted, additional explanation requested, or an alternative task provided. When a rich activity history is available, the BiLSTM model is used to generate the next suggestion; otherwise, heuristics are relied upon.

4 Heuristic Layer of the Recommender Module and Personalization

The recommender module within the Mentora ChatBot system was designed to offer learners content that is most relevant at a given moment, in accordance with their prior interactions, achieved results, and interests. At the present stage of development, a heuristic approach is employed that combines explicit and implicit learner data.

This approach constitutes the first layer of the recommendation system and establishes a foundation for later integration with more complex data-driven models.

4.1 Heuristic Strategies

The heuristic layer operates via a set of rules that map observable behavioral patterns to concrete recommendations. Key rules include:

– **Revisiting areas of difficulty** – if repeated incorrect answers are observed within the same topic, additional lessons and tasks from that domain are recommended;
– **Suggesting the next lesson in sequence** – when a lesson is successfully mastered, the next item in the syllabus sequence is suggested;
– **Accounting for prior interest** – if pronounced interest in a topic is evident (e.g., requests for extra examples or questions posed), content related to that concept is prioritized;
– **Adapting to session length and frequency** – for shorter or less frequent sessions, shorter and easier tasks are proposed to help maintain motivation.

4.2 Personalization via Rules

Although not based on machine learning, these heuristics enable a baseline level of personalization and content adaptation. The rules are currently hand-crafted but were informed by insights gained during the pilot testing phase. In subsequent versions, it is planned that rules be generated dynamically from discovered behavioral patterns.

At present, the rules are implemented directly in the system code; however, a management interface is planned to allow teachers or administrators to edit and tailor them to the needs of specific learner cohorts.

4.3 Integration with the Tutor

This heuristic layer is directly coupled with the system's LLM tutor component. Based on the triggered rules, a contextual prompt is constructed and employed within the dialog. For example:

> "Given that the previous lesson on fractions appeared challenging, it is recommended that another example be worked through together. Would you like to proceed?"

In this way, recommendations are not displayed as a separate list but are naturally woven into the flow of conversation, consistent with the principles of conversational recommender systems [3]. This increases the likelihood that the recommendation will be perceived as supportive guidance rather than an algorithmic command, which positively affects engagement.

4.4 Limitations of the Heuristic Approach

The heuristic approach is valuable for rapid testing and basic personalization; however, limitations arise with more complex modeling of learner behavior, especially when multiple factors interact or when learner needs are highly variable.

In addition, heuristics may be insufficiently robust when a mismatch between measured performance and expressed confidence is present, e.g., when tasks are solved correctly but uncertainty or requests for additional explanation are expressed in dialog.

Despite these limitations, heuristics provide a stable and easily adaptable starting layer for recommendations. Their principal advantages lie in simplicity and transparency. In what follows, their extension via sequential modeling of learner behavior is explained.

5 Sequential Modeling of Learner Behavior Using Logs and a BiLSTM Network

5.1 Log Collection and Processing

Log files constitute the foundation for analyzing user behavior in the Mentora system. This approach enables detailed mapping of the user journey and supports advanced personalized learning. Every interaction with the Mentora ChatBot—questions posed, lesson access, correct and incorrect answers, time-on-task, and answer type (textual, multiple-choice, visual input)—is recorded in pseudonymized form in a relational database.

The data are transformed into sequences that represent learning-interaction flows. Each learning session is encoded as a sequence of attribute vectors (e.g., lesson identifier, task type, performance, response time, topic), thereby enabling the application of sequential prediction models.

Prior research has shown that logs from real educational systems encode behavioral patterns that can be mined to improve recommendations and predict learning outcomes [8,6]. A similar strategy is adopted in Mentora to construct the predictive layer. All logs are stored in accordance with privacy and ethical principles: personal identifiers are pseudonymized, and data access is strictly restricted.

5.2 BiLSTM Model Structure

A bidirectional LSTM (BiLSTM) architecture is employed for sequential modeling of learner behavior, well-suited for uncovering patterns in interaction streams.

The model consumes input sequences composed of event vectors with the following components:

- activity type (lesson, quiz, question),
- topic category (e.g., fractions, grammar, logical reasoning),
- activity outcome (correct/incorrect),
- execution time and inter-event pauses,
- prior recommendations and accepted suggestions.

These vectors are normalized and arranged into fixed-length input tensors for training.

The BiLSTM layer processes the sequence in both forward and backward directions, yielding a contextualized representation at each step. In this manner, it is enabled to account for how preceding and subsequent events influence the current timestep, providing a deeper understanding of learning patterns and more precise next-step predictions. Complex trajectories are thus captured, including performance oscillations, topic transitions, and responses to earlier recommendations.

The implementation is based on the PyTorch framework. The model is planned to be trained on sessions collected during the platform's test phase (August–September 2025), with accuracy and F1-score evaluated on a validation set.

5.3 Predicting Learners' Next Steps

On the basis of the BiLSTM output representations, the Mentora system can generate personalized recommendations that anticipate the learner's most likely next step, including:

- recommending the next lesson consistent with prior performance and topic,
- proposing a task type that best suits the learner (visual, textual, interactive),
- assessing the need for a concept explanation or additional guidance,
- flagging to the instructor a potential misunderstanding.

Such recommendations are injected into the Mentora ChatBot dialog, enabling proactive, intelligent system behavior aligned with the goals of conversational recommender systems in education [3,8]. In this way, the system not only reacts to learner actions but also anticipates needs, creating the experience of an intelligent, supportive tutor. Future development includes automated evaluation of recommendations using engagement and performance metrics, as well as experimental comparisons with static heuristic systems. By integrating the BiLSTM model into the Mentora architecture, the platform transitions from a static, rule-based approach to a dynamic system that tracks—and shapes—the learning trajectory in real time.

6 Results and Discussion

6.1 Evaluation Protocol and Statistical Considerations

The pilot was conducted during February–June 2025, including a seven-day intensive testing window, with $n = 22$ eighth-grade students. Learners were assigned to an experimental (BiLSTM) and a control (heuristics) group ($n = 11/n = 11$). Interactions were logged uniformly across groups. Given the pilot scale, we emphasize effect sizes (absolute deltas) and 95% bootstrap confidence intervals (10,000 resamples) rather than null-hypothesis significance testing. Reported gains (Precision@3 +0.17, F1 +0.19; HR 76%) are interpreted as indicative and consistent across sessions.

Compliance with Ethical Standards. All interactions were pseudonymized, and participation was conducted with parental consent, in line with institutional ethical guidelines.

6.2 Experimental Results

The evaluation outcomes are shown in Figure 3. The BiLSTM-based recommendation module was found to outperform the heuristic approach across all metrics. The largest differences were observed in recommendation accuracy and session duration, indicating a higher degree of personalization and greater learner engagement.

The quantitative results of the pilot study are as follows:

- **Precision@3:** 0.71 (BiLSTM) vs. 0.54 (heuristics);
- **F1 score:** 0.68 (BiLSTM) vs. 0.49;
- **Hit Rate:** 76% (BiLSTM);
- **Average session length:** 12.4 minutes (BiLSTM) vs. 9.8 minutes;
- **User satisfaction:** 4.3 out of 5 (Likert-scale survey).

These findings corroborate the initial hypothesis that sequential modeling enables a deeper understanding of learning trajectories and more accurate prediction of next steps. Beyond improved precision, a higher level of learner engagement was observed, reflected in longer sessions and more frequent interaction with the tutor.

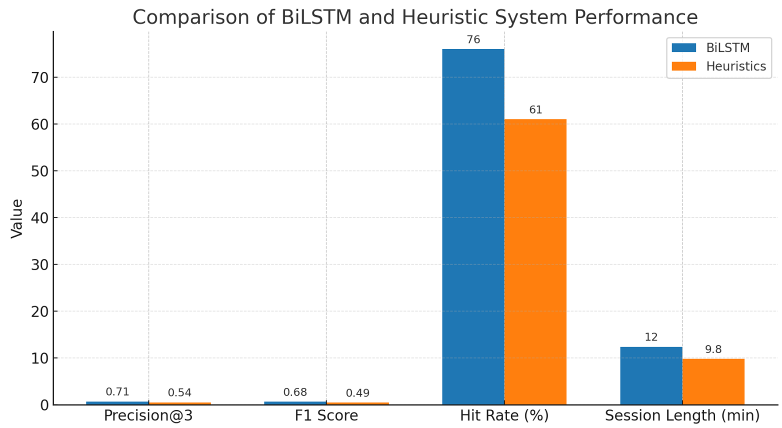

Fig. 3. Performance comparison between the BiLSTM model and the heuristic system

6.3 Discussion and Contributions

The results situate this work within current research on conversational recommender systems (CRS) and sequential modeling in education. In contrast to approaches focused solely on dialog without deep sequential modeling [3], a BiLSTM was integrated to capture temporal dependencies in learning logs. Relative to knowledge tracing that applies LSTM outside the dialog loop [7,6], the contribution lies in embedding the sequential model directly into the dialog flow. Furthermore, by introducing a RAG-based explanation layer, existing efforts on explainable recommendation are extended [13], providing pedagogically meaningful justifications (e.g., "this is a prerequisite for ...", "difficulties were previously observed in ...").

An empirical contribution was established through a pilot evaluation in a real educational setting (22 students, >600 interactions), where measurable gains of the BiLSTM over heuristics were demonstrated (Precision@3 +0.17; F1 +0.19; HR 76%). These results indicate that combining an LLM-driven tutor with a BiLSTM sequence model yields more relevant, adaptive, and explainable recommendations than traditional heuristic approaches.

Limitations. The pilot study was constrained by a relatively small sample and partially synthetic logs. Future work is planned to involve larger cohorts, ablation studies, and enrichment of learner profiles with metadata on motivation, prior knowledge, and learning goals.

7 Conclusion and Future Work

In this paper, the *Mentora ChatBot* was presented—a hybrid conversational recommender system (CRS) for educational settings that combines an LLM-driven tutor, heuristic rules, and a BiLSTM model over learning logs for on-policy decision-making. The system architecture and data flow were described, and a pilot evaluation was conducted under real final-exam preparation conditions (eighth grade), with a focus on recommendation quality and learner acceptance.

Contributions. (i) Integration of *LLM + BiLSTM + RAG* into a single, on-policy decision flow (the LLM manages the dialog, the BiLSTM models sequential context, and the RAG layer provides explanations understandable to learners and teachers); (ii) empirical confirmation that sequential context improves recommendation relevance and uptake over a heuristic baseline (22 students, >600 interactions): *Precision@3* 0.71 vs. 0.54, *F1* 0.68 vs. 0.49, *Hit Rate* 76%; (iii) operationalization of a CRS in a school setting (mathematics and Serbian language) with minimal changes to the existing instructional flow.

Our study has important limitations: a small, convenience sample, irregular study sessions, and partially synthetic logs in the early phase. These factors may bias the results, inflate apparent performance, and materially constrain both internal and external validity; hence, the findings should be viewed as preliminary. In future work we will (i) expand the dataset with real, end-to-end logs from a larger and more diverse student cohort and additional subject domains, (ii) run systematic ablations to isolate component contributions (LLM-only, BiLSTM-only, RAG-only, and combinations), (iii) examine Transformer-based architectures for sequential modeling, and (iv) deepen personalization by learner profile, coupled with richer, pedagogically grounded explanations. To enhance transparency and interpretability, we will include anonymized student–chatbot dialogue snippets in the supplementary material and present additional examples during the conference talk.

In sum, it was demonstrated that the *LLM+BiLSTM+RAG* hybrid can deliver pedagogically meaningful, contextually adaptive, and empirically better-accepted recommendations than static heuristics, laying the groundwork for scalable and transparent learning support in real classroom practice.

Acknowledgments This study was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and these results are parts of the Grant No. 451-03-136/2025-03/200132, with University of Kragujevac - Faculty of Technical Sciences Čačak.

References

1. Drachsler, H., Kalz, M.: Panorama of recommender systems to support learning. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 421–451. Springer (2015). https://doi.org/10.1007/978-1-4899-7637-6_12
2. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Elsevier, 2nd edn. (2006)
3. Jannach, D., Resnick, P., Tuzhilin, A., Zhang, Y.: Conversational recommender systems. Foundations and Trends in Information Retrieval **16**(1), 1–101 (2021). https://doi.org/10.1561/1500000068

4. Liu, B.: Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer, 2nd edn. (2011)
5. Manouselis, N., Drachsler, H., Verbert, K., Duval, E.: Recommender Systems for Learning. SpringerBriefs in Electrical and Computer Engineering, Springer (2012). https://doi.org/10.1007/978-1-4614-4361-2
6. Okubo, F., Yamashita, T., Shimada, A., Ogata, H.: Learning analytics to improve learning outcomes: A case study of a flipped classroom. Research and Practice in Technology Enhanced Learning **12**(1), 1–23 (2017). https://doi.org/10.1186/s41039-017-0043-6
7. Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., Sohl-Dickstein, J.: Deep knowledge tracing. In: Advances in Neural Information Processing Systems (NeurIPS) (2015)
8. Quadrana, M., Cremonesi, P., Jannach, D.: Sequence-aware recommender systems. ACM Computing Surveys **51**(4), 66:1–66:36 (2018). https://doi.org/10.1145/3190616
9. Ricci, F., Rokach, L., Shapira, B.: Introduction to recommender systems handbook. In: Recommender Systems Handbook, pp. 1–35. Springer (2011). https://doi.org/10.1007/978-0-387-85820-3_1
10. Roux, L., Dagorret, P., Etcheverry, P., Nodenot, T., Marquesuzaa, C., Lopisteguy, P.: A multi-layer architecture for an e-learning hybrid recommender system. In: Proceedings of the 18th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2021). pp. 179–187 (2021)
11. Russell, M.A.: Mining the Social Web. O'Reilly Media, 1st edn. (2013)
12. Sun, Y., Papamitsiou, Z., Giannakos, M.: Design and evaluation of a conversational recommender system for learning pathways. IEEE Transactions on Learning Technologies **16**(1), 95–108 (2023). https://doi.org/10.1109/TLT.2022.3185156
13. Zhang, Y., Chen, X.: Explainable recommendation: A survey and new perspectives. Foundations and Trends in Information Retrieval **14**(1), 1–101 (2020). https://doi.org/10.1561/1500000066