

Augmenting AI with Curated Learning Analytics Literature: Building and Initial Exploration of a Local RAG for Supporting Teachers (LARAG)

Sonsoles López-Pernas^{1*}, Ibrahim Belayachi^{1,2, †}, Hesham Ahmed^{1, †}, Ramy Elmoazen^{1, †} and Mohammed Saqr^{1†}

¹ University of Eastern Finland, Yliopistokatu 2, 80100 Joensuu, Finland

² Université de Technologie de Compiègne, R. du docteur Schweitzer CS 60319, 60203 Compiègne Cedex France

Abstract

Though LLMs have completely taken the world by storm, their use in academic settings still faces significant challenges. One of these challenges is that LLMs sometimes “hallucinate” when they do not have the necessary information to reply to the user prompt and, even when they do, they fail to provide a trusted source to back up their claims. In this article, we explore the use of retrieval-augmented generation (RAG) as a way to overcome the aforementioned limitation and enable evidence-based LLM-generated insights. Specifically, we provide the results of our initial exploration of LARAG, a RAG-based system aimed at providing learning analytics recommendations based on the existing literature. Our initial impressions about the system are that it may offer some benefits over traditional LLMs. However, these initial benefits are far from groundbreaking or very accurate.

Keywords

large language models (LLMs), retrieval-augmented generation (RAG), generative artificial intelligence, learning analytics

1. Introduction

Large Language Models (LLMs) are deep-learning algorithms that have been trained on massive amounts of textual data. They contain numerous parameters that enable them to model and predict text [1]. Such characteristics allow LLMs to handle practically any topic ever discussed without any additional training. LLMs have been used in education in diverse ways, for example, as a way to personalized tutoring, ease content creation, and provide automated scoring and feedback [2].

However, among the worries about LLMs' use in education is that LLMs sometimes struggle with factual accuracy [3] and fail to provide a source for the information they present [4]. Researchers have come up with a solution to ensure that LLMs draw their information from a trusted source: Retrieval-Augmented Generation (RAG). RAG consists in supplementing an LLM (Large Language Model), which already has a vast text database for generating text in response to various user requests, with additional files to provide context [5]. This process enables an already powerful LLM to draw on external documents to contextualize and optimize responses, without having to re-train the model each time it is used. Moreover, the responses provided by the RAG will be more reliable and accompanied by sources, unlike LLMs which, due to their obligation to always generate an answer, can sometimes give false information when they do not have the necessary training. In terms of cost, training a RAG is much more advantageous, with a less time-consuming and less computationally intensive process than re-training an LLM [6].

Due to the novelty of RAG, its use in academic settings has been so far scarce with few studies examining its potential. An example is the work by Lee [7], who developed a RAG-based tutor

Proceedings for the 15th International Conference on e-Learning 2024, September 26-27, 2024, Belgrade, Serbia

* Corresponding author.

† These authors contributed equally

✉ sonsoles.lopez@uef.fi (S. López-Pernas); ibrahim.belayachi@etu.utc.fr (I. Belayachi);

hesham.ahmed@uef.fi (H. Ahmed); ramy.elmoazen@uef.fi (R. Elmoazen); mohammed.saqr@uef.fi (M. Saqr)

🆔 0000-0002-9621-1392 (S. López-Pernas); 0000-0002-5792-1340 (R. Elmoazen); 0000-0001-5881-3109 ((M.

Saqr))



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

capable of providing correct and relevant answers regarding R programming. Similarly, Henkel et al. [8] used a RAG-based system to generate responses to students' math questions based on a math textbook. Zhong et al. [9] developed a RAG-augmented generative artificial intelligence agent capable of effectively fostering learners' performance in collaborative problem-solving. Lastly, Yan et al. [10] used RAG to enhance learning analytics dashboards with contextualized explanations.

In this article, we provide an initial look at how the RAG system —augmented with curated up-to-date learning analytics research— can guide teachers to obtain evidence-based insights and recommendations in learning analytics. The system was built at the University of Eastern Finland (UEF) and was named LARAG (referring to Learning Analytics RAG). It is worth noting that this is an initial exploration that aims at demonstrating the system rather than an exhaustive evaluation, which will follow in an extended paper.

2. The RAG process

RAG consists of two main phases: (1) document indexing, and (2) search and answer generation (see Figure 1). In the first phase —document indexing—, the user uploads the curated documents to be indexed by the RAG system. In our cases, we systematically curated all empirical learning analytics research that attempted to predict student performance. These files are typically text-based and should be as structured as possible to aid the next steps of analysis. The input files are then divided into smaller chunks (e.g., paragraphs or sentences). Each chunk is converted into a high-dimensional vector (e.g., 768 or 1024 dimensions) using an embedding model [5], typically based on Transformer architectures (such as BERT). The embeddings capture the semantic content of each chunk, enabling similarity comparison between chunks and user queries. The embeddings are stored in a vector database, optimized for similarity search. Some examples of vector databases are Pinecone, Weaviate, Faiss, and Milvus [11]. The vector database supports efficient search algorithms such as the Approximate Nearest Neighbor (ANN) search, which finds vectors closest to the query vector. Metadata (e.g., source document ID or section) is often stored alongside each vector, allowing to trace back the source of the retrieved chunks at the time of answer generation.

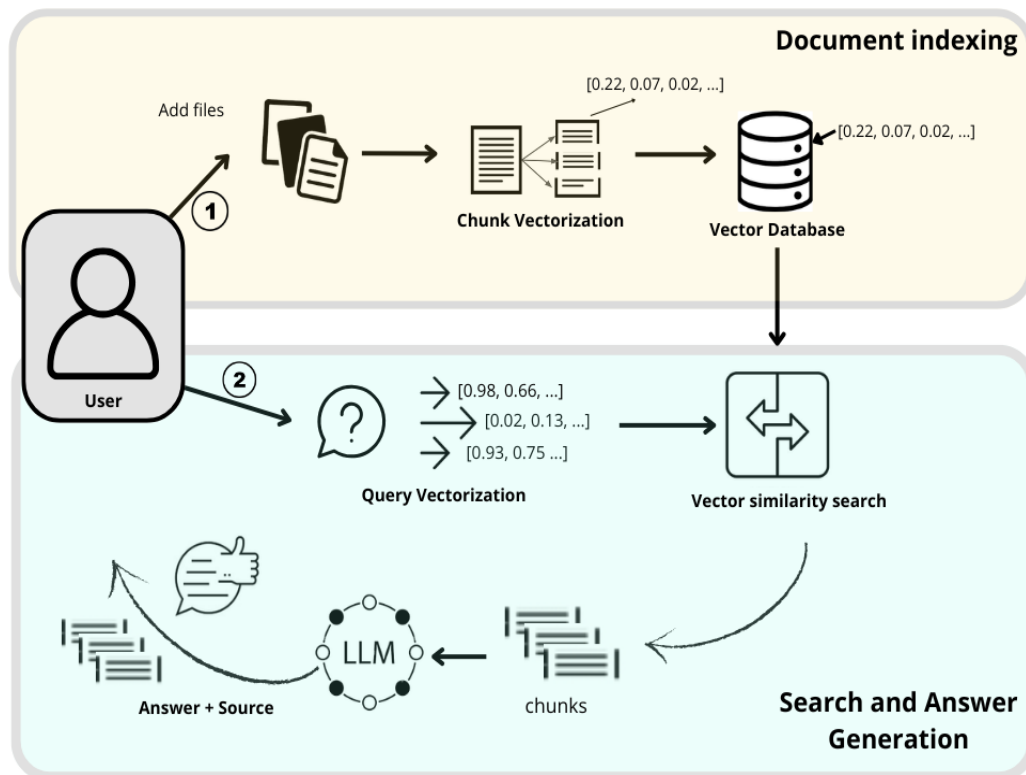


Figure 1: RAG process

The second phase of RAG —search and answer generation— starts by the user inputting a query —i.e., prompt—, similar to interacting with any LLM. The query is vectorized using the same embedding model that was used on the document chunks so that the query's vector representation

is in the same semantic space as the chunk vectors, thus, enabling comparison. The query vector is matched against chunk vectors in the vector database using similarity measures like cosine similarity or dot product. These metrics calculate the closeness of vectors based on the angle or magnitude in vector space. The search algorithm efficiently finds the top k closest vectors (relevant chunks) to the query vector, returning the most contextually relevant document parts. The retrieved chunks (top k closest matches) are provided to the LLM for answer synthesis. These chunks are often concatenated or structured for the LLM to interpret them as context.

The LLM generates a response by using the query and relevant chunks as context. The model synthesizes an answer based on both the question and the retrieved information, which results in accurate and contextually appropriate responses. The final output includes the answer synthesized by the LLM, often along with citations or references to the original documents. Thus, using RAGs, as opposed to only LLMs, increases transparency and trust, since users know where the factual information came from.

3. Initial Look at LARAG to Explore How It Generates Evidence-based Insights in Learning analytics

In this section, we describe our initial results regarding the suitability of using LARAG to generate evidence-based insights following the latest learning analytics literature. The purpose is to enable teachers, administrators, and even students themselves —i.e., those who are not familiar with research practices and results— to obtain rigorous recommendations based on existing literature. In this way, we hope to contribute to bridging the gap between learning analytics research and practice.

To achieve this goal, we used an open-source RAG system —Kotaemon [12]— that supports both phases of the RAG process: the document upload, and the querying of the documents using a chat-like format. Kotaemon is built with Python and integrates technologies like LangChain for orchestrating the RAG pipeline; ChromaDB as a vector database to store and retrieve document embeddings, and supports LLMs from multiple sources such as OpenAI, Azure or local (open source) LLMs such as Ollama. Moreover, Kotaemon provides a web user interface based on Gradio that allows both the uploading of documents and the interaction with them through chat.

After installing and deploying Kotaemon, we uploaded a corpus of learning analytics articles using Kotaemon's web interface. Specifically, we provided the full text of 136 articles that cover the topic of predictive learning analytics since this is one of the central themes of learning analytics that is of interest to researchers and practitioners alike. To find all relevant articles about this topic, we first downloaded all the related literature reviews using the following query: (*predictive AND learning AND analytics AND systematic AND literature AND review*), limited to after 2011, and journal and conference papers. This step resulted in 13 reviews. We obtained the references of each review, which amounted to 1517 (1427 after duplicate removal). We screened all studies and included only those that predict a target variable (e.g., engagement, achievement, dropout), contain the word *predict** in the title, abstract, or keywords, rely on student data, and create a predictive model to predict the target variable. As a final result, 136 papers were included and their respective PDF files were downloaded and loaded to LARAG.

Once the documents are loaded, the LARAG system is ready to respond to related questions. We chose Ollama as our LLM since it is an open-source solution that achieves very high accuracy at the level of commercial solutions. Figure 2 shows the web interface of LARAG. The left side of the interface allows choosing which documents are used to retrieve the responses. The middle part contains the conversation prompts and responses. The right side showcases the documents that were used to generate the response.

To evaluate the suitability of LARAG to provide evidence-based insights compared to a usual LLM (without context augmentation), we used the same query to prompt our system with and without RAG. Table 1 shows the prompt and the results with and without context augmentation from the uploaded articles.

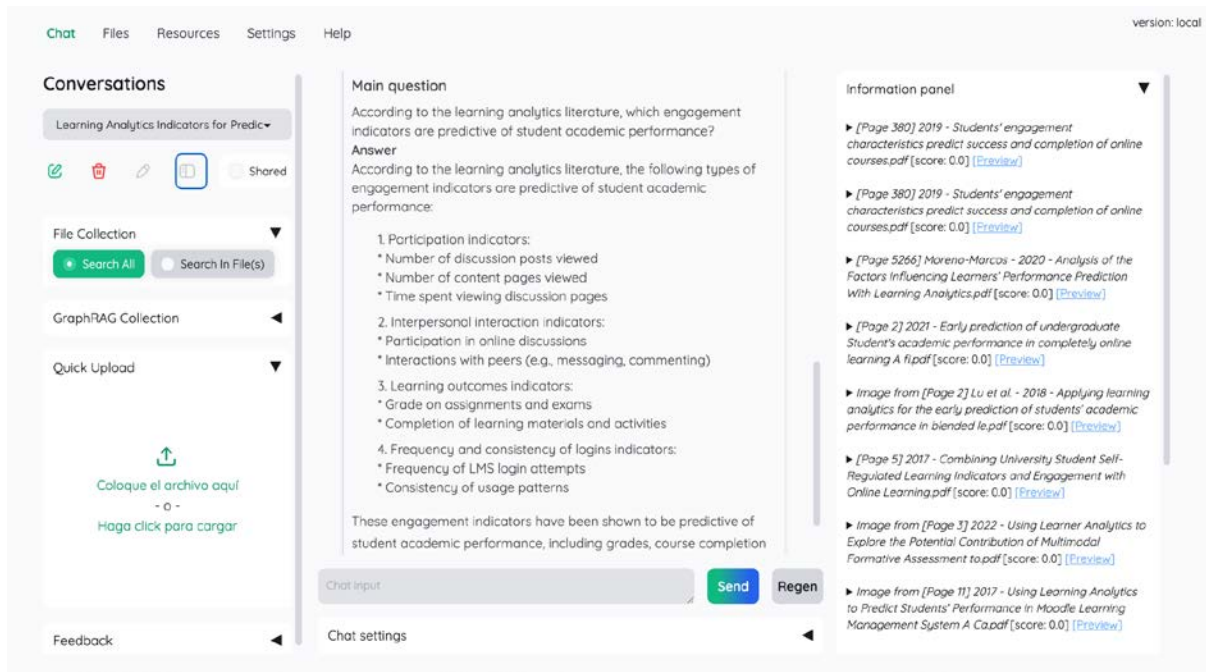


Figure 2: LARAG web interface provided by Kotaemon

Table 1
Prompt and results with and without RAG

Prompt	
<p>According to the learning analytics literature, which engagement indicators are predictive of student academic performance?</p>	
Without RAG	With RAG
<p>According to the learning analytics literature, several engagement indicators have been found to predict student academic performance. These include:</p> <ol style="list-style-type: none"> Classroom participation: Studies have shown that students who actively participate in class discussions and activities tend to perform better academically. Homework completion rates: Students who regularly complete their homework assignments tend to demonstrate higher levels of academic success. Online activity metrics: Engagement with online learning platforms, such as logging in frequently or completing online activities, has been linked to improved academic performance. Self-reported engagement measures: Students' self-reports of their motivation and engagement have been found to be strong predictors of academic success. Screen time and session duration: Research has suggested that students who spend more time on specific learning platforms or screens tend to demonstrate better academic performance. Quiz and exam scores: Consistently high scores on quizzes and exams are robust indicators of student engagement and academic achievement. Social presence metrics: Measures of students' social interactions and engagement with peers through online platforms have been linked to improved academic outcomes. <p>It is essential to note that the specific engagement indicators used can vary depending on the context, discipline, and population being studied.</p>	<p>According to the learning analytics literature, the following types of engagement indicators are predictive of student academic performance:</p> <ol style="list-style-type: none"> Participation indicators: <ul style="list-style-type: none"> * Number of discussion posts viewed * Number of content pages viewed * Time spent viewing discussion pages Interpersonal interaction indicators: <ul style="list-style-type: none"> * Participation in online discussions * Interactions with peers (e.g., messaging, commenting) Learning outcomes indicators: <ul style="list-style-type: none"> * Grade on assignments and exams * Completion of learning materials and activities Frequency and consistency of logins indicators: <ul style="list-style-type: none"> * Frequency of LMS login attempts * Consistency of usage patterns <p>These engagement indicators have been shown to be predictive of student academic performance, including grades, course completion rates, and overall success in online courses.</p>

We evaluated the quality of both responses based on the following criteria: *accuracy*, *specificity*, *justification*, *usefulness*, *implementability*, *coherence*, and *equity*. These criteria are based on different user-centric evaluation frameworks used to evaluate recommendation systems in general and

conversational-based ones specifically [13]–[17]. *Accuracy* refers to how aligned is the output to the desired request; *specificity* refers to how precise and specific there response is; *justification* refers to the provision of an explanation as to why the output was provided; *usefulness* denotes the utility of the output; *implementability* refers to the ease of which a recommendation is applied, *coherence* refers to the level of the consistency between the different parts of the output, and *equity* describes the fairness in resources and support distribution to account for imbalances and differences.

Based on these criteria, it seems that both responses point to comparably *accurate* engagement metrics that are known to be predictive of student academic achievement. The RAG response uses terminology that is more aligned with learning analytics research, whereas the non-RAG response uses general concepts that may be more understandable to a general audience. As such, the RAG response contains much more *specific* metrics to represent similar constructs to what the non-RAG response proposed. For instance, the response without RAG pointed to “social presence” broadly as a relevant engagement indicator, while the RAG-based response went one step further and proposed specific quantitative metrics: “number of discussion posts viewed”, or “time spent viewing discussion pages”.

Both responses *justify* the results by broadly declaring “according to the learning analytics literature”, although they may be simply mimicking the user prompt. An advantage of RAG is that we can see the actual sources used to generate the RAG response (see Figure 2), whereas we cannot access that information for the non-RAG response. However, although the RAG points to the part of the paper that was used to generate the response, the RAG did not provide a direct citation for each engagement metric pointing to which articles were consulted to generate that specific recommendation.

Regarding *usefulness*, the non-RAG response contains more general –or shall we say generic– indicators that can be applied broadly to more learning contexts whereas the RAG response was tailored to online learning (even though that was not specified in the prompt). Nonetheless, the RAG response points to specific quantitative indicators that have been empirically proven to be associated with student achievement in the specific context of online or blended learning. In turn, the non-RAG response includes more broad metrics that may or may not be associated with performance depending on their operationalization. Perhaps for a non-expert audience the non-RAG response is more useful to understand what the engagement indicators that need to be collected represent, though the RAG response points to the specific operationalization thereof.

In terms of *implementability*, the RAG response suggested mostly engagement metrics that can be easily derived from unobtrusive data collection through the Learning Management System (LMS). The non-RAG response is mostly unclear in terms of how to calculate the metrics; although some also involve using LMS data, others are more costly, such as using self-reports (there is no specific instrument suggested for this).

When it comes to *coherence*, both responses contain significant overlap in the metrics suggested. For instance, the non-RAG response suggested measuring “online activity” and “session duration”, whereas one could argue that session duration is a measure of online activity. The RAG response is better structured in categories and related metrics, although some of these categories somewhat overlap, such as “participation indicators” and “interpersonal interaction indicators”. However, neither of the two responses included contradictory recommendations and both were mostly coherent.

Lastly, regarding *equity*, neither of the two responses explicitly mentions the possible bias in this data collection or the existence of individual differences that need to be taken into account. However, the non-RAG response did point out that “the specific engagement indicators used can vary depending on the context, discipline, and population being studied” highlighting that these indicators do not generalize to all contexts, which is one of the main findings of learning analytics literature on performance prediction [18]–[20].

In summary, the RAG responses are of higher quality overall, providing more accurate, implementable, and structured insights relevant to learning analytics. The non-RAG response might appeal to a broader audience but lacks precision and implementability in an educational context.

4. Reflection and Future Directions

In this exploratory paper, we introduced our initial thoughts on LARAG, a local RAG system built with curated learning analytics research at the University of Eastern Finland. The system is baked

on open-source software that is easy to build and customize. The curated learning analytics research was highly relevant and represents the current state of the art. Our initial impressions about the system are that it may offer some benefits over traditional LLMs, namely more specific and easy-to-implement recommendations. However, these initial benefits are far from groundbreaking or very accurate. For instance, although the RAG points to the specific chunks that were used to generate the response, it does not map specific parts of the response to these chunks (i.e., citations).

In the future, we aim to offer a systematic evaluation across several research domains and relevant questions and evaluate the system using a rigorous methodology. The LARAG system described is our initial build and is still in an experimental stage. Future systems are planned to have far larger datasets that contain massive educational datasets.

Acknowledgments

This article has been co-funded by the European Commission through the **ISILA** (2023-1-FI01-KA220-HED-000159757) project.

Declaration on Generative AI

During the preparation of this work, the author(s) used Ollama, Koteamon for Table 1 at the page 8. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] M. McDonough, "Large language model," *Encyclopedia Britannica*, 22-Jan-2024. [Online]. Available: <https://www.britannica.com/topic/large-language-model>. [Accessed: 05-Nov-2024].
- [2] E. Kasneci *et al.*, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learn. Individ. Differ.*, vol. 103, no. 102274, p. 102274, Apr. 2023.
- [3] I. Augenstein *et al.*, "Factuality challenges in the era of large language models and opportunities for fact-checking," *Nat. Mach. Intell.*, vol. 6, no. 8, pp. 852–863, Aug. 2024.
- [4] D. Peskoff and B. Stewart, "Credible without credit: Domain experts assess generative language models," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Toronto, Canada, 2023, pp. 427–438.
- [5] P. Lewis *et al.*, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.
- [6] H. Soudani, E. Kanoulas, and F. Hasibi, "Fine tuning vs. Retrieval Augmented Generation for less popular knowledge," *arXiv [cs.CL]*, 03-Mar-2024.
- [7] Y. Lee, "Developing a computer-based tutor utilizing Generative Artificial Intelligence (GAI) and Retrieval-Augmented Generation (RAG)," *Educ. Inf. Technol.*, pp. 1–22, Nov. 2024.
- [8] O. Henkel, Z. Levonian, C. Li, and M. Postle, "Retrieval-augmented generation to improve math question-answering: Trade-offs between groundedness and human preference," *Proceedings of the 17th International Conference on Educational Data Mining*. International Educational Data Mining Society, pp. 315–320, 2024.
- [9] X. Zhong, H. Xin, W. Li, Z. Zhan, and M.-H. Cheng, "The Design and application of RAG-based conversational agents for collaborative problem solving," in *Proceedings of the 2024 9th International Conference on Distance Education and Learning*, Guangzhou China, 2024, pp. 62–68.
- [10] L. Yan *et al.*, "VizChat: Enhancing learning analytics dashboards with contextualised explanations using multimodal generative AI chatbots," in *Lecture Notes in Computer Science*, Cham: Springer Nature Switzerland, 2024, pp. 180–193.
- [11] P. N. Singh, S. Talasila, and S. V. Banakar, "Analyzing embedding models for embedding vectors in vector databases," in *2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG)*, Indore, India, 2023, pp. 1–7.
- [12] Cinnamon, *kotaemon: An open-source RAG-based tool for chatting with your documents*. Github, 2024.
- [13] P. Pu, L. Chen, and R. Hu, "A user-centric evaluation framework for recommender systems," in *Proceedings of the fifth ACM conference on Recommender systems*, Chicago Illinois USA, 2011.

- [14] H. Kunstmann, J. Ollier, J. Persson, and F. von Wangenheim, "EventChat: Implementation and user-centric evaluation of a large language model-driven conversational recommender system for exploring leisure events in an SME context," *arXiv [cs.IR]*, 05-Jul-2024.
- [15] Y. Jin, L. Chen, W. Cai, and X. Zhao, "CRS-Que : A user-centric evaluation framework for conversational recommender systems," *ACM Trans. Recomm. Syst.*, Nov. 2023.
- [16] P. Jurado de Los Santos, A.-J. Moreno-Guerrero, J.-A. Marín-Marín, and R. Soler Costa, "The term equity in education: A literature review with scientific mapping in web of science," *Int. J. Environ. Res. Public Health*, vol. 17, no. 10, p. 3526, May 2020.
- [17] D. Jannach, "Evaluating conversational recommender systems: A landscape of research," *Artif. Intell. Rev.*, vol. 56, no. 3, pp. 2365–2400, Mar. 2023.
- [18] R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat, "Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS," *IEEE Trans. Learn. Technol.*, vol. 10, no. 1, pp. 17–29, Jan. 2017.
- [19] M. Saqr, J. Jovanović, O. Viberg, and D. Gašević, "Is there order in the mess? A single paper meta-analysis approach to identification of predictors of success in learning analytics," *Studies in Higher Education*, vol. 47, no. 12, pp. 2370–2391, Dec. 2022.
- [20] J. Jovanovic, S. López-Pernas, and M. Saqr, "Predictive modelling in learning analytics: A machine learning approach in R," in *Learning Analytics Methods and Tutorials*, Cham: Springer Nature Switzerland, 2024, pp. 197–229.