Recommendation Systems Education using Project-Based Learning (PBL)

Victor Takashi Hayashi[1], **Ivan Fellipy Gonçalves Ferreira**[1]

[1]Instituto de Tecnologia e Liderança (Inteli)
Prof. Almeida Prado, 520 – 05508-070 – São Paulo – SP – Brazil

victor.hayashi@prof.inteli.edu.br, ivan.ferreira@sou.inteli.edu.br

Abstract. *Recommendation systems are a crucial component of digital platforms, enabling personalized user experiences across industries. This study explores the integration of Project-Based Learning (PBL) in teaching recommendation systems to undergraduate students, a method aimed at enhancing their understanding of this complex Artificial Intelligence (AI) application through real-world scenarios. Through PBL, students engaged in collaborative, client-focused projects that simulated real industry challenges, promoting technical skills and problem-solving competencies essential for computing careers. The methodology combined quantitative and qualitative analyses: student performance data was tracked across multiple project sprints to evaluate skill progression, and in-depth interviews with faculty were conducted to gather insights on PBL's efficacy in recommendation system education. The results demonstrate that PBL not only boosts technical proficiency in developing recommendation models but also reinforces teamwork and critical thinking skills, providing a framework for effective AI education in higher learning. This case study with 80 participants contributes valuable findings to the adoption of active learning methods for preparing learners for real-world AI challenges.*

1. Introduction

Recommendation systems are a technology broadly used to suggest content to users based on their behavior, past activities, likes and dislikes. It is present in lots of today's online services, such as e-commerce, streaming services, social networks, food delivery and more. Fundamentally, recommendation systems' operation is related to Artificial Intelligence (AI) techniques and algorithms, which are trained with a huge amount of data to, through probabilistic calculations, recommend content that would likely be within users interests. Due to its growth as a relevant, meaningful and powerful application of AI - and almost 20 years after the initial development of automated collaborative filtering [Konstan et al. 2014] - recommendation systems gradually became a thing to be formally taught by education institutions, especially in courses and programs related to software engineering.

It is important to note, however, that some of these courses and programs run by means of teaching methodologies that do not satisfy contemporary demand of indistinctly including all public, which is brought by social and technological transformation of the last decades [Trindade and Souza 2023]. Such traditional methodologies, unlike active learning methods like Problem-Based Learning, do not benefit learning itself and practical knowledge application, as concluded [Firmo et al. 2023].

Incidentally, project-based learning (PBL) is a strategic student-centered methodology. Through PBL methodology, students can master the required technical knowledge and skills while diving in real problems. The main benefits of this teaching strategy are associated with the practical and sometimes ludic activities that drive students to a better understanding of concepts involved in the problem solution. Moreover, when paired with agile methodologies, it can help students develop collaborative working, not to mention it also contributes to developing skills like holistic vision, critical thinking and solving problems [dos Santos et al. 2021]. In computing education, this problem-based approach sets an environment that promotes both technical and nontechnical skills development [dos Santos et al. 2021].

Thus, "How does PBL methodology contribute to recommendation systems teaching?" and "How does PBL methodology impact missing data scenarios in recommendation systems?" are the main questions that motivate this study. We describe a case study with Project-Based Learning with industry partners (Dell and SEDUC) for Recommendation Systems Education of two Software Engineering classes (n=80).

2. Related Work

This section presents scientific research about the use of active learning methodologies, as well as presenting how the present work aims to contribute to the field. Only works related to using Project-Based Learning or Problem-Based Learning to teach Artificial Intelligence (AI), or some Brazilian study cases using PBL in similar fields.

The first study, conducted by Genard Trindade, reports a teaching experience in training thirty undergraduate Computer Science students in the subject of Artificial Intelligence Applied to Education. Active teaching strategies were employed to create a collaborative environment, encouraging debates on teaching AI in Basic Education. The study reveals that the analysis of the students' evaluation processes revealed satisfactory learning perceptions. The course presented was conducted at the Center for Higher Studies of Lábrea, University of the State of Amazonas, focusing on active methodological approaches to foster discussions on AI and Machine Learning, aiming to prepare future teachers to integrate these technologies into education[Trindade and Souza 2023].

In another study, André Firmo describes, in his study about using Robocode to teach AI subjects to undergraduate students, how the use of an active methodology combined with gamification can enhance student inclusion and engagement. The author details that developing a combat robot allowed students to build interdisciplinary skills and competencies. His research aimed to investigate the benefits of using the Robocode tool in the teaching-learning process. The study involved 79 students from an AI course and was conducted through a descriptive empirical quantitative research approach. Its findings indicate an increase in attendance and academic performance, especially among students with greater learning difficulties[Firmo et al. 2023].

Aside that, some scientific research published in the Brazilian Symposium on Informatics in Education (SBIE) has significantly been a reference for this work. For instance, Alice Finger describes her implementation of Problem-Based Learning in teaching programming paradigms at the Federal University of Pampa. Her results indicate improved student autonomy and performance, particularly in procedural paradigms, though challenges remained with logical and functional paradigms[Finger et al. 2021].

Moreover, Lidiany Santos also explores the application of PBL in her study, specifically in an Applied Health Informatics course at the Federal University of Sergipe. The findings suggest that PBL facilitated significant student engagement and practical skills development through the creation of mobile health applications. Her research highlights the effectiveness of PBL in promoting practical skills and engagement, which is directly relevant to the current study's focus on recommendation systems[Santos et al. 2020].

However, as stated by André Firmo [Firmo et al. 2023], brazilian research of this nature - involving undergraduate students performances through active methodologies - are not satisfactory present in the literature. Alongside that, there are even fewer study cases about this process with recommendation systems teaching, which represents a gap in the literature considering recommendation systems are an increasing AI application being taught in schools nowadays.

Considering that, the present work was done to contribute to this gap on study cases about the teaching-process of recommendation systems through active methodologies.

3. Methodology

The course division and application of the PBL model in this study follow the approach outlined by [Vargas et al. 2020], who emphasizes the importance of student cooperation in complex environments to develop both learning abilities and professional skills. Vargas' method involves small student work groups, typically without the constant presence of a tutor, where tasks are realistic, professional, and of significant magnitude. These tasks often cover only a small portion of the course content, encouraging students to deepen their technical knowledge independently and focus on group-oriented product development.

In our study, students were similarly divided into groups of six to eight members. Each group was tasked with solving a real-world problem related to recommendation systems. The groups operated autonomously, with the instructor's role being primarily that of a facilitator, guiding the workgroups as needed. The methodology promoted self-management, teamwork, leadership, and communication skills among the students, aligning with the active learning approach emphasized by [Vargas et al. 2020].

The data collection for that was conducted through the institution's web platform, where all activities are managed. This platform acts as a central hub for students to access and complete both individual and group tasks, as well as to review the theoretical content covered throughout the course. The platform records the grades achieved by each student and group for every activity. It also allows exporting data in spreadsheet format, facilitating a comprehensive analysis of student performance and engagement throughout the course.

Furthermore, exploring the collected data was possible because of Pandas library in Python, which provided robust tools for data manipulation and analysis. By leveraging Pandas, it was possible to efficiently clean, filter, and organize the data extracted from the platform, which included analyzing student performance metrics, identifying patterns in group work dynamics, and evaluating the impact of different activities on learning outcomes.

The dataset extracted from the platform included the following columns: 'turma', 'projeto', 'uuid-aluno', 'grupo', 'activity-id', 'atividade', 'descricao-da-atividade', 'enunciado', 'barema', 'link', 'tipo-de-atividade', 'semana', and 'nota'. Using the Pandas library, the 'uuid-aluno' column was removed since it contained unique values for each row and did not correlate with other variables relevant to the analysis. Additionally, duplicate entries were eliminated using the drop-duplicates() method to ensure that each record was distinct. The next step involved filtering records where the 'nota' column contained a value of -1, which indicated invalid or missing scores. Following this, the data was segmented by activity, creating smaller dataframes for each specific activity type. Each of these smaller dataframes was further divided by group, allowing for detailed analysis of each group's performance in comparison to the activities completed.

Alongside that, another library used was matplotlib. This library was employed to generate visualizations of the groups' progress. Specifically, group grades were plotted to compare their performance over sprints. For this, we used Pandas to create DataFrames containing the grades for each group, as well as the overall class average for each sprint.

When considering the outcomes of the analysis enabled by Pandas, it became clear that the course was structured into weekly activities to ensure continuous peer learning and iterative project development. These activities were categorized into five key categories:

1. Self-study, in which students engaged in independent study about topics that were going to be discussed in class. For example, a self-study about non-relational databases [Silva et al. 2021];
2. Instructional meetings, a class that is not primarily expository, but a discussion about topics that students had already seen in self-study activities and how those topics could be applied in the project context. For example, instructional meetings about non-relational databases and how these are related to the project's demands;
3. Project development, where students collaboratively worked on their projects, applying theoretical concepts to real-world problems, like bringing up the system architecture based on the interactions with the non-relational database;
4. Guidance meetings, sessions where the instructor provided feedback, addressed challenges, and ensured the projects were progressing effectively;
5. Assessment/research, where students conducted evaluations and research. For example, evaluating the other members of their work group.

This structured approach was implemented in two distinct projects from the Inteli Recommendation Systems module. The first project, conducted by Class 6, which did a partnership with SEDUC, focused on addressing inefficiencies in the control and monitoring of material and supply deliveries to state schools in São Paulo. The solution involved developing a mobile application for tracking and managing deliveries, integrated with BIG DATA management to optimize data governance in the delivery process.

The second project, carried out by Class 3 alongside Dell Company, aimed to improve talent retention by creating a mobile application that matches projects with employees for Dell Company. This system was designed to increase employee engagement and motivation by offering content that encourages participation in company projects and learning new topics. Both projects utilized recommendation systems to enhance the decision-making processes within their respective domains.

In addition to the quantitative analysis of student performance, a complementary qualitative approach was also employed to gather insights into the effectiveness of the Problem-Based Learning (PBL) methodology in teaching recommendation systems. An interview was conducted with the professors responsible for running the recommendation systems module. The interview aimed to capture their considerations on the impact of PBL on student learning, the challenges faced during the course, and their perceptions of how well students assimilated the key concepts of recommendation systems. The interview was semi-structured, allowing for flexibility in the conversation while ensuring that key topics were covered. The audio of the interview was recorded and transcribed using Whisper, a speech recognition and transcription model developed by OpenAI, implemented in Python.

4. Results

The dataset extracted from the institutional platform enabled a thorough analysis of student performance throughout the course, especially in relation to group dynamics and progression across the five sprints. This analysis provides key insights into how the Problem-Based Learning (PBL) methodology influenced the development of technical and non-technical skills in the context of recommendation systems.

The analysis focuses on group performance, highlighting how each group's project grades evolved during the sprints. The comparison between each group's median grades and the overall class median allows for an in-depth understanding of the problem-solving strategies adopted by different teams and their progression over time. This is crucial for evaluating the practical application of PBL in teaching recommendation systems.

For example, Figure 1 shows the median project grades of each group across all five sprints, plotted alongside the median project grades of the entire class. This visual comparison makes it possible to assess each group's performance in two key ways: first, how the group performed relative to itself over the course of the sprints, and second, how the group's progress compared to the other groups and the class as a whole.

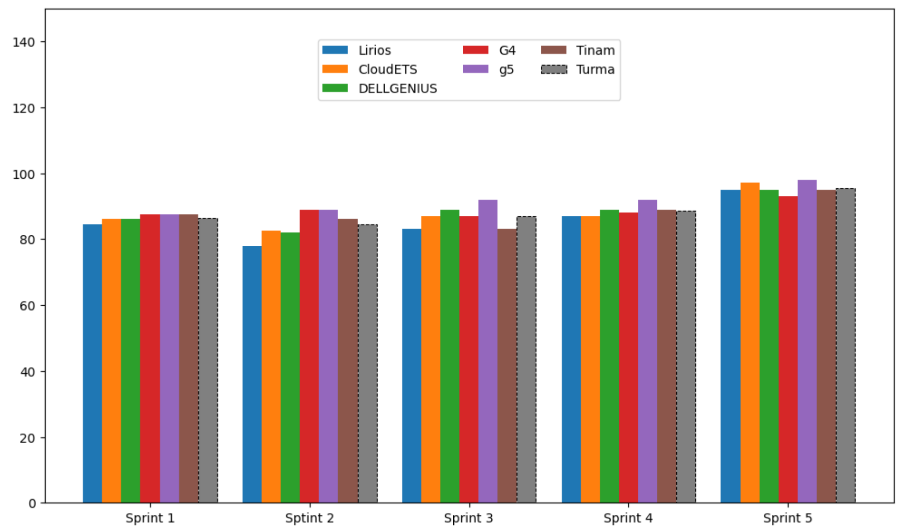

Figure 1. Groups' project scores medians through sprints

For instance, Figure 2 illustrates the median test scores of each group across all test intervals. This comparison enables two main types of assessment: first, it allows us to observe how each group's test performance evolved over time, providing insights into their consistency or improvement across the module; second, it facilitates a comparison between each group's median test scores and other groups' median, indicating whether groups scored above or below another group average at each interval.

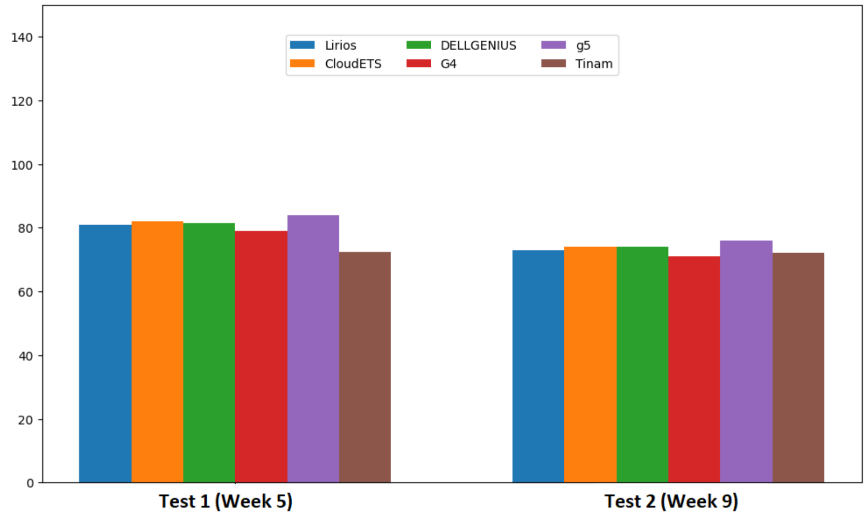

Figure 2. Groups' tests scores medians

Figure 3 provides a broader view of the trends in project scores over the entire module. The trend line shows the general progression of median project scores from the start to the end of the module, offering insight into the trajectory of each groups' performance. This visualization helps identify whether there was a trend of improvement, stability, or decline in project grades over time, reflecting the overall impact of the module's progression on group projects.

By examining these trends, it is possible to identify whether specific groups struggled or excelled at different stages of the project-based tasks. Moreover, this comparison highlights how well individual teams adapted to the iterative problem-solving tasks inherent to the PBL methodology, revealing their growth in both collaborative and technical competencies.

Beyond quantitative analysis, qualitative insights from two professors involved in the recommendation systems module were gathered to deepen understanding of the Problem-Based Learning (PBL) methodology's impact. One professor supervised students' projects, while the other provided technical instruction in programming. Their perspectives highlighted both the advantages and challenges of using PBL in this context, emphasizing its role in fostering student engagement, problem-solving adaptability, and a hands-on grasp of recommendation systems concepts. The professors' reflections, captured in a semi-structured interview transcribed using Whisper by OpenAI, revealed nuanced observations on students' skill development and readiness for practical AI applications, further supporting the potential of PBL in preparing students for real-world scenarios.

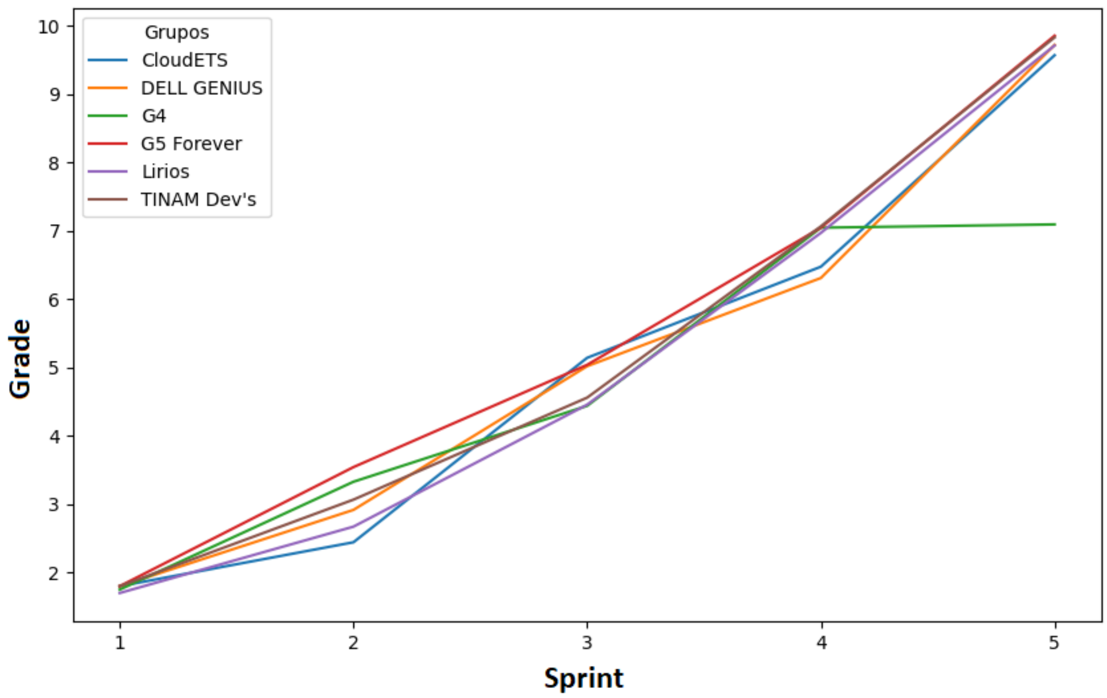

Figure 3. Groups' tests scores tendency

5. Analysis

The analysis of student performance focused on evaluating group dynamics, skill acquisition, and the effectiveness of the PBL methodology in teaching recommendation systems. Quantitative insights were derived from performance metrics tracked across five sprints. Each sprint represented an iterative stage in project development, fostering a continuous learning process.

The performance scores across different groups highlighted progressive improvements in technical and collaborative competencies. A consistent increase in median grades was observed from Sprint 1 to Sprint 5, suggesting that iterative project development under PBL was effective in consolidating both technical knowledge and teamwork capabilities. For instance, groups with lower initial scores demonstrated significant recovery over subsequent sprints, attributed to the structured feedback and guidance sessions.

Qualitative analysis, derived from professor interviews, further corroborated these findings. Professors noted that students were better equipped to apply theoretical concepts to practical scenarios, particularly in modeling and implementing recommendation systems. However, they also emphasized challenges such as uneven participation among group members and the steep learning curve associated with advanced AI concepts. Furthermore, professors expressed skepticism about any direct correlation between strict adherence to agile methodologies and improved group performance, highlighting the need for further investigation.

While the analysis primarily focused on the recommendation systems module, comparisons between the two projects—Class 6's supply chain optimization for SEDUC and Class 3's talent-matching platform for Dell—illustrated the versatility of PBL. Both projects showcased how contextually distinct real-world problems can foster comparable learning outcomes, including critical thinking and problem-solving. Despite differences in technical focus, both classes achieved high levels of engagement and creativity, validated by their final project deliverables. However, as no direct data analysis of the SEDUC project was conducted, these observations are based on qualitative feedback.

Visualizations created using matplotlib provided additional insights into sprint-wise performance variations. These plots highlighted consistent trends, such as the positive impact of instructional and guidance meetings on subsequent project outcomes.

The application of Project-Based Learning (PBL) to the teaching of recommendation systems revealed several valuable lessons, both in terms of pedagogical strategies and practical aspects.

From a pedagogical standpoint, the iterative nature of PBL facilitated progressive learning and skill acquisition. By structuring the course into sprints and incorporating real-world challenges, students demonstrated increasing autonomy, critical thinking, and collaboration. Instructional and guidance meetings played a crucial role in consolidating knowledge, as evidenced by improved project outcomes in subsequent sprints. This confirms the value of combining theory-driven content with hands-on, contextualized application.

In terms of technical skill development, students showed an improved ability to translate theoretical AI concepts into functional systems. However, the course revealed a limitation in time allocation for more complex topics. Teachers indicated that these concepts were sometimes too advanced for the course's duration, suggesting a need for curricular adjustments or preparatory activities to support student learning more effectively.

Regarding teamwork and soft skills, the study highlighted the diversity of group dynamics. While some teams exhibited steady progress, others faced challenges related to uneven participation. This finding underscores the importance of implementing strategies to monitor and balance individual contributions within collaborative environments.

The inclusion of external partners such as Dell and SEDUC significantly enriched the learning experience by providing continuous feedback. Students responded positively to real-world problems, which increased engagement. Nevertheless, the study suggests that involving partners earlier in the planning phase could enhance the alignment between educational goals and practical outcomes, maximizing the impact of these collaborations.

6. Conclusions

This study presented how Project-Based Learning (PBL) can be used when teaching recommendation systems, emphasizing the integration of real-world challenges to foster technical and non-technical skills. The iterative nature of PBL facilitated cumulative learning, as evidenced by improvements in group performance and student engagement over sprints. The structured feedback and guidance sessions were crucial in addressing the challenges students faced in understanding complex concepts such as similarity metrics and recommendation algorithms.

However, the findings also highlight areas for improvement. Both professors emphasized the need for better time allocation to cover complex topics, such as constructing recommendation functions, which were perceived as too advanced for the allotted schedule. This limitation underscores the necessity of revising the module's structure to allow more comprehensive exploration of these critical concepts. Additionally, the potential for "unplugged" activities—simplified, non-digital exercises to introduce recommendation concepts—was identified as a promising pedagogical strategy for reducing cognitive load and enhancing understanding.

The qualitative insights also suggest that the success of PBL relies heavily on student engagement and intrinsic motivation rather than strict adherence to methodological principles. This observation opens avenues for further research to explore how different student profiles impact learning outcomes in PBL environments and how educators can tailor their approaches to maximize inclusivity and effectiveness.

Future work could also investigate how to better integrate partners into the early stages of project planning, ensuring alignment between educational objectives and real-world applications. Moreover, expanding this study to other computing domains or different educational settings could provide a broader understanding of PBL's applicability and scalability.

By addressing these challenges and pursuing these opportunities, educators and researchers can continue to refine PBL methodologies, contributing to more effective and inclusive teaching strategies in computing education.

References

dos Santos, S. C., Reis, P. B. S., Reis, J. F. S., and Tavares, F. (2021). Two decades of pbl in teaching computing: A systematic mapping study. *IEEE Transactions on Education*, 64(3):233–244.

Finger, A., da Silva, J. P., and Ecar, M. (2021). Utilizando aprendizado baseado em problemas para o ensino de paradigmas de programação. In *Anais do XXXII Simpósio Brasileiro de Informática na Educação*, pages 135–144, Porto Alegre, RS, Brasil. SBC.

Firmo, A., Abreu, B., Chagas, M., Júnior, J. S., and Rodrigues, L. (2023). Uso do robocode como ferramenta de ensino de inteligência artificial nos cursos de computação. In *Anais Estendidos do XXII Simpósio Brasileiro de Jogos e Entretenimento Digital*, pages 1429–1434, Porto Alegre, RS, Brasil. SBC.

Konstan, J., Walker, J., Brooks, D., Brown, K., and Ekstrand, M. (2014). Teaching recommender systems at large scale: Evaluation and lessons learned from a hybrid mooc. volume 22, pages 61–70.

Santos, L., Neves, D., and Menezes, F. (2020). An assessment of project-based learning towards applied health informatics teaching in an undergraduate class. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 202–211, Porto Alegre, RS, Brasil. SBC.

Silva, L. F. C., Riva, A. D., Rosa, G. A., et al. (2021). *Banco de Dados Não Relacional*. SAGAH, Porto Alegre. E-book, p. Capa.

Trindade, G. and Souza, D. (2023). Inteligência artificial aplicada à educação: Um relato de experiência docente na formação de acadêmicos de licenciatura em computação. In

Anais do XXIX Workshop de Informática na Escola, pages 843–854, Porto Alegre, RS, Brasil. SBC.

Vargas, E., Alfaro, M., Fuertes, G., Lillo, S., Ternero, R., Sabattin, J., Banguera, L., Durán, C., and Muller, M. (2020). A project based learning approach for teaching artificial intelligence to undergraduate students*. *International Journal of Engineering Education Vol. 36, No. 6, pp. 1773–1782, 2020.*