JM0130-M-6

DATA CONSULTANCY IN ACTION

GROUP 1

30/05/2025

Safeskip: Optimizing Student Learning Paths

Data-Driven Adaptivity on a higher Level

Consultants

Name	Student Number	Email Adress
Guido Beijer	snr 2061356	g.beijer@tilburguniversity.edu
Bram Donkers	snr 2042954	b.j.t.donkers@tilburguniversity.edu
Collin Leppink	snr 2065734	c.j.a.leppink@tilburguniversity.edu
Lars van Rijckevorsel	snr 2065745	l.p.a.vanrijckevorsel@tilburguniversity.edu
Bram Zegger	snr 2044065	b.w.zegger@tilburguniversity.edu
Quinten van Halewijn	snr 2083234	q.d.vanhalewijn@tilburguniversity.edu

Contents

1 Excecutive Summary

This report summarizes the results of a consulting project that us JADS students under DCiA Group 1 carried out in collaboration with Gynzy, a top educational tech provider in Belgium and the Netherlands. The purpose was to research and develop a personalized learning path recommender system called *SafeSkip* that could advise when students might safely skip learning objectives (also known as ”microgoals”) in Gynzy’s hierarchical curriculum without sacrificing the quality of their education.

At the moment, Gynzy’s adaptation is limited to Elo scoring-driven exercise-level changes. This approach works well for adjusting to immediate challenges, but it doesn’t help students navigate more extensive paths of learning or find unwanted objectives they may skip. The consulting team’s goal was to use the huge dataset, which included over 4.2 billion student interactions and over 2,500 learning goals, to provide higher-level personalization.

A six-sprint, Scrum-based approach was used to complete the project, and Gynzy and academic supervisors provided ongoing feedback. The initial stages were devoted to the theoretical investigation of relevant modelling techniques, such as SVD, PCA, MIRT, and Dynamic Bayesian Networks. In the end, ALS (Alternating Least Squares) was chosen due to its ability to handle sparse data and adapt to collaborative filtering needs.

To predict ELO scores on unattempted microgoals, the ALS model was trained using historical data and a time-based train-test split. Predicted ELO scores over a safety threshold of 1.0 were considered ”good skips.” With a recall of 85 percent, the model showed the ability to identify students who were prepared to skip. Nevertheless, its precision of 42 percent raised concerns over the possibility of false positives. In order to prevent logically incorrect recommendations, hierarchical limits are used to limit skip recommendations to only subjects that have already been explored.

The skip logic was further improved in response to the input received during Sprints 4 and 5 to maintain the structure of the curriculum and to match the educational standards. In the final deliverables are a technical report, a functional Python prototype, and a consultative presentation with specific implementation advice.

Important suggestions are to include temporal dynamics into ALS (recurrent models, time-weighted updates), via dependency mapping to expand the scope beyond previously addressed subjects, and beginning pilot testing to assess system usability and teacher trust in the real-life classroom. The initiative also stressed the importance of open communication with teachers, suggesting that the *SafeSkip* results should be used to complement human judgement rather than replace it.

The results show that introducing intelligent skip logic to Gynzy’s learning platform is feasible, setting the stage for more personalized and effective learning. *SafeSkip* is an important first step towards more adaptivity in primary education, even though advancements, particularly in precision, are required for actual adoption.

2 Introduction & Background

2.1 Policy frameworks and operational context

Policy frameworks

National and international educational policy efforts are emphasizing on the need for personalized education. The goal of the Dutch government's ongoing investments in digital education changes is to use technological innovations to better meet the different needs of students. According to European Commission. Directorate General for Education and Culture [2] at the European level encourages the development of adaptive learning technology and the need for inclusive, high-quality digital learning environments as well.

In the Economic Co-operation and Development [1] publication from OECD the use of AI systems in education is stimulated encourages to improve teaching in responsible, intelligent ways, personalizing learning, and giving students more control. Taken together, these frameworks emphasize how important it is to personalize learning paths for each student, which makes intelligent systems like the one suggested for Gynzy both relevant and consistent with more general goals of educational reform.

Operational context

Gynzy is a Dutch EdTech startup that serves 200,000 students and 8,200 schools in Belgium and the Netherlands. A Teacher Platform and a Student Learning Platform (SLP) make up the platform they use. At the moment, the SLP uses exercise difficulty according to student performance using an Elo-based ranking system. This adaptability, though, is limited to the individual exercise level.

Gynzy is investigating intelligent curriculum-level adaptivity as it sees the possibility of progressing beyond this. Finding out if students may safely skip specific "microgoals", which are learning objectives within a hierarchical curriculum structure, without sacrificing mastery is the specific goal. This would improve student engagement and learning efficiency, which would be in line with their overarching goal of providing efficient, individualised education at scale.

2.2 Motivation of consult project

This consulting work was driven primarily by the need to move from exercise-level adaptivity to learning goal-level personalisation. Gynzy's current system makes sure that students receive exercises that are suitable for their current level of performance, but it doesn't tell them which learning objectives they may safely ignore or which they should work on next.

Gynzy is in a good position to benefit from innovative modelling techniques because it has access to more than 4.2 billion student responses and a curriculum with more than 2,500 microgoals. Finding valid skip opportunities, however, calls for more than just effective research. It also calls for predictive, personalised knowledge based on curriculum structure and student ability.

Our group got the responsibility to come up with and adopt a solution to this problem. The final goal was to suggest a data-driven approach that would enable students to skip over microgoals they have successfully achieved, allowing them to customise their curriculum journey without sacrificing level of learning.

2.3 Consultancy terms of reference

Goal:

The purpose of the consultation was to help Gynzy develop a personalised, scalable learning path recommender that could make safe skip recommendations at the microgoal level. The partnership aimed to use Gynzy's historical student data to verify the feasibility and effect of such a system.

Scope:

- Using past data to design a skip-recommendation system.
- Evaluating the system's performance and educational validity.
- Providing a technical report and a functional prototype.
- Providing useful recommendations for future implementation.
- Integration with the real Gynzy platform and complete system was not covered by the project.

Key tasks:

- Using Gynzy's datasets to exploratory data analysis.
- Building hybrid recommendation models, such as PCA, MIRT, and ALS.
- Defining definitions for good and bad skip logic.
- Comparing predictions to time-based splits of actual data.
- Using stakeholder input to continuously enhance the model.

Deliverables:

The technical report provides explanation of the problem's background, the analytical approach, and the models' functionality. The working prototype shows how the skip recommendation system works by making predictions based on past student data. Last but not least, Gynzy received advice via the consultation summary presentation.

2.4 Project methodology

We used a data-driven, Scrum-based approach that fitted to Gynzy's technical and educational requirements. The team explored, developed, and validated a proof of concept for personalized learning route recommendations in six iterative sprints lasting two weeks each. Our work was guided by regular feedback sessions with Gynzy and academic supervisors, which also made sure that it was in line with educational standards.

The first stages were devoted to theoretical investigation. We looked into a number of modelling techniques, including Dynamic Bayesian Networks, PCA, and MIRT, to see how well they might represent the ability of learners and direct adaptive learning. This made it easier to choose suitable modelling approaches without committing to implementation too soon.

We used Multidimensional Item Response Theory (MIRT) and Alternating Least Squares (ALS) for analyzing historical student data throughout the modelling phase. MIRT helped probability prediction of micro-goal mastery, while ALS enabled collaborative filtering on sparse datasets. Model performance was benchmarked using a time-based train-test split, with precision serving as the main test parameter to reduce suggestions for bad skips.

Based on feedback from stakeholders, we improved our skip logic as the project went on, defining what a "safe skip" is and setting limits to avoid making irrational suggestions. We summarised our results, verified the prototype, and created specific implementation suggestions for Gynzy during the last sprint.

3 Factual/Contextual Information

3.1 Projectgoal

As mentioned in the introduction, the goal is to predict when a student can safely skip a microgoal within the hierarchical structure of Gynzy. Before these skips can be predicted, it must first be defined when a skip is considered safe. This is done using the ELO score with a threshold of 1.0. A good skip occurs when the student has an ELO score greater than or equal to 1.0, and a bad skip occurs when the score is below 1.0.

3.2 Data

For this project, the dataset provided by Gynzy was used. As described in Section 3.1, a student is considered to have mastered a microgoal if the ELO score is greater than or equal to 1.0. To structure this information clearly, a matrix $R \in \mathbb{R}^{n \times m}$ was created, where n is the number of students, m the number of microgoals, and R represents the ELO scores. After this, the data was split using a time-based train-test split, making it possible to evaluate the model on how well it predicts future ELO scores.

3.3 Model

Based on the earlier definition of when a skip is considered correct, it becomes possible to build the model. The final model used for collaborative filtering is an Alternating Least Squares (ALS) model. This model was chosen because it is well-suited for sparse data, which is the case here since students have only completed a small portion of all available microgoals. ALS works in this scenario by factorizing the original matrix R into two smaller matrices: U and V. Matrix U contains latent factors for the students, and matrix V contains latent factors for the microgoals. By taking the dot product of these two matrices, a prediction can be made for the ELO score of a student on a microgoal they have not yet completed. The formula for ALS is:

$$\hat{R}_{ij} = U_i \cdot V_j^T$$

where:

- $\hat{R}_{ij}$ Predicted ELO-score of student i on microgoal j
- U_i Latent factor student i
- V_j^T Latent factor microgoal j

To prevent students from taking shortcuts in the learning material, the ELO score is only predicted for worlds or islands where the student has already completed at least one microgoal. This approach supports a more structured, personalized, and pedagogically sound learning experience by guiding students to deepen their understanding within familiar topics before moving on to new ones. If the output of the model is $\hat{R}_{ij}$ >= 1.0 the model considers it safe for student i to skip microgoal j.

3.4 Evaluation

To properly evaluate the performance of the ALS model, a baseline model must first be defined, since Gynzy does not yet use an implementation like this. This baseline predicts, for each microgoal, the average ELO score of all students who have previously completed it. Just like with the ALS model, these predicted ELO values are only filled in when the student has already completed at least one microgoal within the same world or island. The reason for filling in the missing data this way in the baseline is that these predictions are completely non-personalized.

Next, the ALS model is evaluated using 3-fold cross-validation, and two types of evaluation metrics are considered. First, the Root Mean Squared Error (RMSE) is used, as it gives an indication of how accurately the model predicts the ELO score. The formula for RMSE is:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{(i,j)\in n} (R_{ij} - \hat{R}_{ij})^2}$$

Where:

- N Amount of test data.
- R_{ij} Real ELO-score of student i on microgoal j.
- $\hat{R}_{ij}$ Predicted ELO-score of student i on microgoal j.

	Student passed	Student failed
Model predicts skip	True positive (good skip)	False positive (bad skip)
Model predicts no skip	False negative (missed opportunity)	True negative (correct caution)

Figure 1: Confusion matrix

In addition, the model's precision is evaluated using the confusion matrix. The confusion matrix can be seen in Figure 1.

The confusion matrix shows the four possible outcomes. In this project, it is especially important to minimize the number of false positives (bad skips). This means the model recommends skipping a microgoal even though the student has not yet mastered the content, which can negatively affect the student's learning process. For this reason, precision is considered the most important evaluation metric. However, evaluating precision is difficult to calculate in a real implementation, since once a microgoal is actually skipped, it is no longer possible to know what the outcome would have been if the student had completed the task. This means that counterfactual terms are used for evaluation.

4 Findings

The final product developed in this project, *SafeSkip*, was specifically targeted at improving the educational experience of primary school students using Gynzy's learning platform. The usability of the model was evaluated not only in terms of algorithmic performance, but also from the perspective of how well it aligns with the needs of students, teachers, and Gynzy's development team. By tailoring recommendations to previously visited topic areas, it ensures educational consistency while helping students avoid redundant tasks. This has the potential to boost engagement and motivation among learners by maintaining an optimal level of challenge.

4.1 The road towards ALS collaborative filtering

During our search for the best solution, we initially explored several candidate models that could strike a balance between maintaining educational integrity and leveraging data science techniques. The first modeling phase included a comparative exploration of Bayesian Knowledge Tracing, Dynamic Bayesian Networks, and Multidimensional Item Response Theory (MIRT). While Bayesian approaches offered probabilistic tracking of student understanding over time, they often lacked scalability and interpretability in sparse datasets. MIRT, on the other hand, provided a principled framework for modeling latent student abilities across multiple dimensions and was better suited to the structure and sparsity of our data. Thus, it became the conceptual foundation for our subsequent research.

In the second modeling phase, we sought practical implementations of MIRT-like behavior through matrix factorization techniques. Here, we compared Principal Component Analysis (PCA), Singular Value Decomposition (SVD), and Alternating Least Squares (ALS). PCA and SVD offered strong dimensionality reduction capabilities, but lacked flexibility in handling missing data and regularization in a collaborative filtering context. ALS emerged as the most appropriate model because of its robustness in sparse settings, its ability to factorize the student-microgoal matrix with missing entries, and its suitability for iterative optimization. In particular, ALS can be viewed as an operational counterpart to MIRT, approximating the latent ability and item difficulty parameters in a way that enables personalized predictions at scale. This conceptual alignment between MIRT and ALS justified our selection of ALS as the engine of *SafeSkip*.

4.2 Initial results

In relation to the terms of reference, the system meets the defined objective of making safe skip recommendations at the microgoal level using collaborative filtering techniques. The ALS model demonstrated close to acceptable predictive accuracy with an overall classification accuracy of 70% and a notably high recall of 85%, indicating a strong capability to identify students who are ready to skip. However, a lower precision, which was determined as the most important metric, of 42% highlights the need to carefully consider model outputs, as a significant portion of the recommended skips might still be risky. For this reason, a hierarchical filter was added to reduce the riskiness of this lower precision. This is a short preventative measure; however, this hard limit could be removed when *safeskip* is able to take into account time and therefore learning development throughout a student's career.

The model's evaluation included a comprehensive analysis of a confusion matrix, as well as a quantitative comparison based on RMSE. The cross-validated Root Mean Squared Error (RMSE) achieved by the ALS model was 0.5609 for 10 factors and 10 iterations of the ALS, which is significantly lower than the baseline RMSE of 0.7903 obtained through mean imputation. This represents a 29.0% improvement in predictive accuracy and demonstrates the value of ALS in modeling student-microgoal interactions even in sparse datasets. Although this comparison is somewhat useless as the baseline in production is currently better measured by tracking how a student progresses it does provide some insight in whether the model itself provide any value at all. The confusion matrix below is somewhat more explanatory on whether the skips are safe.

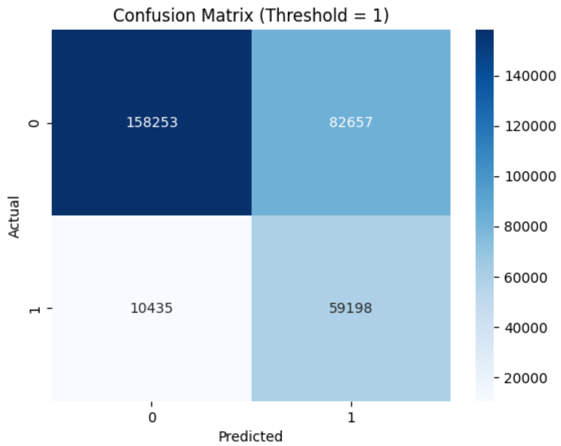

Figure 2: confusion matrix

above we can see the confusion matrix,revealing that 46.2% of the predictions were true negatives, 17.3% were true positives, 23.6% were false positives, and 3.0% were false negatives, based on the total number of evaluated cases. This shows the model's tendency to over-recommend skips in certain contexts, possibly leading to premature progression for some students. While the high recall reduces the likelihood of missing students who are ready to advance, the false positive rate suggests that model recommendations should be treated as advisory and ideally supplemented with teacher oversight or a increased skip threshold (currently 1.0).

Insight into individual predictions further illustrates the model's strengths and limitations. The most accurate predictions corresponded exactly with actual ELO scores, as seen in several examples where predicted and true values matched to two decimal places. At the same time, the largest prediction errors exceeded absolute differences of 6 to 10 points, with certain ELO predictions diverging drastically from actual performance. These outliers likely stem from the sparse and somewhat under optimized dataset and point to a need for model refinements, including mechanisms to handle uncertainty and prevent overconfident predictions. The main takeaway is that the ALS model is able to deal with sparse datasets but due to the static nature of the current implementation is has a hard time ”finding” the meaningful skips. As we will also discuss in the recommendations, these findings highlight several directions for continued development. Integrating temporal dynamics into the ALS model could enhance its ability to reflect evolving student ability over time. Ways of integrating such a temporal term would most likely involve developing some sort of sequence or adding a bias towards goals that occur earlier in a student's learning path.

5 Recommendations

This report has presented a data-driven solution to one of Gynzy's biggest challenges: maintaining student engagement and learning efficiency in a system at risk of being perceived as “one-size-fits-all”. The proposed solution, SafeSkip, aims to introduce skip logic into Gynzy's learning environment. By enabling students to bypass redundant microgoals once they have demonstrated mastery of the underlying knowledge, the system gives a more efficient and personalized learning journey.

5.1 Integrate Temporal Dynamics into ALS Modeling

Alternating Least Squares (ALS) has proven effective in handling sparse data environments, particularly for student-microgoal interactions. However, a key limitation is its static nature. Learning is not static students' abilities change over time, influenced by progression, regression, and stabilization in understanding.

To address this, the ALS model should be enhanced with temporal awareness. One promising approach is to weight interactions based on recency, thereby giving more influence to recent student performance. Another possibility is the adoption of sequential modeling techniques, such as recurrent neural networks or analyses using rolling windows, which are better suited to capture time-dependent patterns. Incorporating decay factors can also help reduce the impact of outdated data.

These adjustments would enable the model to more accurately reflect students' current capabilities and learning trajectories, thereby increasing the precision and contextual relevance of skip logic recommendations.

5.2 Expand Recommendation Scope Beyond Touched Islands

In its current form, the SafeSkip system only offers skip recommendations within content areas that a student has already engaged with. While this constraint ensures structural continuity, it also limits the personalization potential of the platform.

To broaden the system's scope while preserving pedagogical soundness, it is advisable to begin mapping conceptual dependencies across different islands. This would allow the system to identify safe opportunities for cross-island skips. A gating mechanism should be implemented to assess whether a student is conceptually ready for such recommendations, ensuring that any skip remains appropriate and beneficial. Furthermore, piloting this approach across different age groups could help identify which cohorts are most suited for broader skip permissions.

This expansion would allow Gynzy to offer a more dynamic and ability driven learning experience, aligning educational content more closely with student capability rather than rigid curriculum pathways.

5.3 Deployment

With a prototype that is technically sound, the next step is to launch a pilot using the prototype in selected schools. This pilot should evaluate the integration within Gynzy's interface in real-world conditions. Then, feedback from the teachers can be collected regarding trust, interpretability, and override functionalities. With this information, establishing a continuous feedback loop is essential. This looks to refine the system over time. For this monitoring, key performance indicators like student engagement, learning outcomes, and pacing is crucial.

5.4 Strategic and educational implications

Personalized education is not just a technical goal, it is essential for meaningful student growth. By implementing SafeSkip, Gynzy can change from static content to a more responsive, learner-centered platform. Students will benefit from a learning path that is optimized and with that makes the studens feel challenged, while giving educators actionable insights. Over time, this could position Gynzy as a frontrunner in adaptive education.

References

[1] Organisation for Economic Co-operation and Development. **Trustworthy artificial intelligence (AI) in education**. 218. OECD Publishing, Apr. 2020. DOI: 10.1787/a6c90fa9-en. URL: https://www.oecd.org/en/publications/trustworthy-artificial-intelligence-ai-in-education_a6c90fa9-en.html.

[2] Sport European Commission. Directorate General for Education Youth and Culture. **Digital education action plan 2021–2027: Improving the provision of digital skills in education and training**. Luxembourg: Publications Office, 2023. URL: https://data.europa.eu/doi/10.2766/149764.

6 Appendix

6.1 Data Ethics Canvas

Data sources The data that is used in this project was provided by Gynzy. This dataset consist of information. However the data does not contain any form of sensitive information about students, or their school since all data was anonymized.

Right around data source The data was collected and shared by Gynzy for general educational use for this project. Gynzy retains all rights of the data. As a team we signed a NDA to be able to use the data.

Limitations in data source We did not investigate it but it is possbile that the dataset is biased. The dataset may not represent all types of students. Also, the ELO-score of the students are only updated twice a day at fixed times and not directly after completing a microgoal.

Ethical and legislative context The biggest ethical risks in this project are recommending bad skips that negatively affect learning and lack of transparency in decision-making. To prevent this risks we let the model allow skip if the confidence is high enough, which is at an ELO-score higher or equal than 1. We also advise Gynzy if the implement this model that they tell their clients that teachers have to use it as an advice.

Your reason for using data The purpose of this data is to test if collobrative filtering in the form of a ALS model could improve the current version of Gynzy. The purpose of the ALS model is to predict when students can safely skip microgoals which makes it possible to make learning more efficient and personalized.

Positive effect on people Primary school students and teachers will benefit of this solution. Since students get a more personalized learning traject and teachers get more insight in their students. Also Gynzy will benefit from this solution since their platform is able to give more effective and personalized learning.

Negative effect on people This solution can also have negative effect on students when they are allowed to skip a microgoal even tough they do not understand it completely. This can lead to knowledge gaps. Also, their data has to completely anonomyzed to prevent any risk of using their personal data.

When the model is implicated there is also a risk of bias affecting how opportunities are distributed. By this, the model should be monitored closely.

Minimizing negative impact We are minimizing the negative effect by using precision as evalaution. We want the precision to be as high as possible since this prevents that the model has the minimum amount of bad skips. For now, we used an threshold of ELO ¿= 1.0, but if this leads to a low precision it is also possible to set this threshold higher to reduce the amount of skips which gives the model higher confidence. Besides this, we advise Gynzy that if they implement it they should explain their clients that the output only has to be used as an advise and not to

Engaging with people Gynzy was our primairy point of contact since we are doing this project for them. We did not think on how people can appeal or request changes to the product since the product is not live. The goal was to check if ALS could improve the current version of Gynzy and if it could be possible to implement.

Communicating your purpose WIthin the project we had open communication with Gynzy since they want to implement a certain form of collaborative filtering. Since they have most knowledge of their platform and we about data science technics. We had multiple meetings during the project to keep our vision the same. We did not ensure yet how teachers or kid could understand our solution since the objective was to test if ALS was able to improve the current version of Gynzy.

Openness and transparency The methodology , model and results can be openly shared of this project. Only the data is owned by Gynzy. Besides this, in this course we had multiple supervision sessions with other groups that were working on collaborative filtering. In this sessions we could openly show our progress and give feedback to each other. By this, we had a high transparency with other groups.

Sharing data with others Gynzy is the owner of the data, so we are not allowed to share anything about the data with others. They only shared the data with us to allow us to do academic research on collaborative filtering.

Ongoing implementation Currently we did not focus on implementing. The goal is to find a collaborative filtering technique that could improve the current version of Gynzy. However when implementing Gynzy should tell their clients that this model has to be used as an advise for teachers. Also, the model have to give the teacher insights why it is advising that the student could skip a certain microgoal.

Reviews and iterations It is important that if the model is implemented it is monitored often to make sure that the skips are still correctly.

Your actions We recommend launching a pilot phase in selected schools to collect feedback from teachers on model trust. Also there has to made an explanation tool that helps the teacher understand why the model advises that a student can skip the microgoal.