Sensitivity Analysis and Temporal Stability of Student Success Predictors based on Different Data Sources in Education

Marija Pokos Lukinec[1][0009-0006-1564-5970], Dijana Oreški[2] [0000-0002-3820-0126], and Dino Vlahek[3][0000-0002-3911-8685]

[1, 2, 3] University of Zagreb, Faculty of Organization and Informatics, Pavlinska 2, 42000, Varaždin, Croatia
mapokos@foi.hr, dijana.oreski@foi.hr, dvlahek@foi.hr

Abstract. Student success prediction is a central topic in educational data mining and learning analytics, as institutions increasingly rely on data-driven approaches to enhance learning outcomes. However, the dynamic nature of educational environments raises questions about the long-term reliability of predictive features used in these models**.** This study aims to investigate the temporal stability of features extracted by sensitivity analysis of predictive models developed by integrating data from various sources, including the e-learning system, student attendance records, teacher opinions, and meteorological data. In this study, the stability of success predictors is modeled using machine learning algorithms – Random Forest and Gradient Boosted Decision Tree. By applying regression metrics, the precision of the model is assessed to determine the reliability of predictive features over time. Identification of the relevant success predictors and their temporal stability provides insights into significant success predictors in the long term. The results support the development of robust predictive models and highlight key features that contribute to the reliable analysis of student success outcomes.

Keywords: Predictive Data Modeling, Machine Learning, Stability of Success Predictors, Learning Management System, Learning Analytics, Random Forest, Gradient Boosted Tree.

1 Introduction

Machine learning algorithms and the vast amounts of data generated daily have led to data-driven predictions aimed at advancement and development across various domains, including education. To provide high-quality feedback in the shortest possible time, new models and methods for tracking student progress are being developed. Data collected and analyzed from the Learning Management System (LMS) helps predict which features influence student dropout rates and which contribute to motivation, success, and continued engagement. The analysis of data from the e-learning system is defined as learning analytics. According to [1], *learning analytics is the collection, analysis, interpretation and communication of data about learners and their learning that provides theoretically relevant and actionable insights to enhance learning and teaching*. Researchers aim to extract high-quality

Research Paper
DOI: https://doi.org/10.46793/eLearning2025.058PL
Part of ISBN: 978-86-89755-37-4

data, specifically e-learning system features that would serve as a solid foundation for developing accurate and reliable student success models. Study [2] states that only LMS data is insufficient for prediction, highlighting the need for integration, making it one of the research objectives.

This study will analyze data extracted from the e-learning system, integrated with multiple sources, to determine the features that influence the student success model and which features demonstrate stability in prediction over time, thereby ensuring model robustness.

Objectives of the study are: (i) to integrate data from various sources (e-learning system, nastava.foi.hr, teachers' opinion, meteorological data), (ii) to determine the temporal stability of features over time, (iii) to compare the stability of predictive models obtained through different machine learning algorithms, (iv) to identify stable and relevant features from the formative assessment group.

Based on the established objectives, research questions are formulated: (i) Which group of integrated features contributes more to model stability?, (ii) Which success predictors can be extracted as stable within courses in the e-learning system for several years?, (iii) Which stable data features have greater predictive power in predictive models obtained using different machine learning algorithms, based on sensitivity analysis?, (iv) Which features from the formative assessment group have been identified as stable?

2 Course and Data Description

This section describes course included in the research as well as data used for predictive models development.

2.1 Course Description

The research is conducted within a blended course model. The course is delivered at the Faculty of Organization and Informatics, University of Zagreb, at the undergraduate level. It is a mandatory course that enrolls over 200 students annually. The course lasts 15 weeks and is organized so that the resources needed for lectures, exercises, and seminars are published in the e-learning system.

Materials are published in the system during both synchronous and asynchronous classes. Materials used during synchronous classes remain accessible for students to review and study later, while materials published during asynchronous classes additionally serve as preparation tools and support work in a live environment. Since this is a blended course model, in addition to data on student activity within the system and the number of points accumulated through various assignments during the course, data on student attendance records and teacher assessments are also available for analysis. This teaching format enables teachers to closely monitor student progress.

Assignments are given to students either synchronously or asynchronously, and as previously mentioned, all materials are published within the e-learning system. Assessment is conducted through formative (assignments) and summative (midterms, projects) tasks. Students must achieve at least 50% in each summative component to pass the course. The final grade is determined by summing the points earned in summative assessments. Assignments are submitted by students via the e-learning system, reviewed by teachers, and discussed through verbal feedback during live sessions. This process provides students with insights into their performance and areas for improvement.

For the 2023/2024 course, 241 students were enrolled, while in the 2024/2025 course, 246 students were enrolled.

The next subsection presents the data extracted from the e-learning system for a regression task, specifically predicting the number of points obtained in midterms.

2.2 **Data Description**

The data includes integrated information from: (i) the Moodle e-learning system; (ii) student attendance records; (iii) subjective teachers' opinions regarding students' performance during classes, and (iv) meteorological data retrieved from the Open-Meteo website [3].

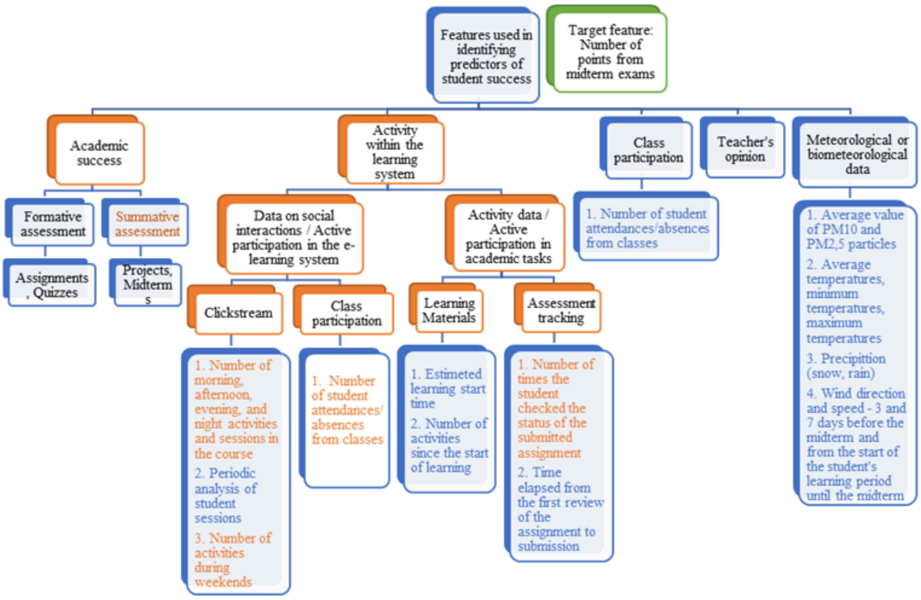

Fig. 1. Feature groups and examples of features used in the research

The first set of 46 features relates to LMS data. Since [4] states that the time spent within a course is not significant for the outcome unless dedicated to learning, the research analyzes those features that can track time spent on learning with the integration of new data. Some of these features, highlighted in orange, were also used in studies [2], [5], and [6].

The second group of features, meteorological features, includes data on average, minimum, and maximum temperature, precipitation, wind speed, and PM particles. PM10, and PM2.5 - airborne pollutant particles capable of penetrating the skin, bronchi, and bloodstream, thus posing a risk to human health [7]. Given their impact on the environment, climate, and visibility [7], it is hypothesized that they also influence human mood to some extent, potentially affecting motivation and academic success. The inclusion of meteorological data further contributes to this research, as there are no existing studies that have integrated such data into their analyses.

Meteorological data were incorporated to explore whether environmental conditions—such as temperature or daylight duration—affect student engagement and overall success. External factors of this kind can influence students both directly and indirectly. For example, severe weather conditions may impact attendance, limit students' mobility, or affect their concentration and motivation. Similarly, prolonged periods of poor weather could reduce opportunities for social interaction and extracurricular learning, while favorable conditions may foster more consistent participation. By including meteorological data, the study investigates whether academic performance is sensitive to non-academic variables, offering a broader view of factors shaping student outcomes.

Attendance record features are entered by teachers and can be retrieved from the nastava.foi.hr system or the e-learning system. The features provide data on the number of times a student has attended classes (lectures and/or seminars, and/or exercises).

The teacher's opinion feature represents the teacher's subjective assessment, based on the student's performance during classes or exercises, as well as their engagement in all aspects of academic work.

Teacher assessments of student engagement and performance, while valuable, introduce an inherently subjective dimension into the predictive model. Such judgments may capture behavioral student patterns that are not reflected in quantitative data from LMS logs or environmental records. However, reliance on teacher opinions must be interpreted cautiously, as individual perceptions can vary widely and may be influenced by implicit biases or limited classroom interactions. Including this predictor provides an opportunity to compare human-informed insights with objective metrics, helping to assess its relative stability and predictive power.

Integration of these data sources serves as the basis for development of predictive models for student success prediction.

3 Research Methods

The research follows the CRISP-DM standard for data analysis. This standard is applied through six phases: domain understanding, data understanding, data preparation, modelling, evaluation, and deployment [8]. In conducting the research, data is first collected from the e-learning system and other sources, then integrated to ensure a comprehensive analysis.

After collecting and explaining the data for understanding, the data is extracted, structured, and prepared for modelling using a machine learning algorithm.

Sensitivity analysis identifies stable features for making predictions. In the evaluation phase, models are assessed, and research questions are addressed.

Data sets are collected at the Faculty of Organization and Informatics, University of Zagreb, and approval from the Ethics Committee is obtained for conducting the research and using the data, with the necessary anonymization of personal indicators and proper storage. The data is divided into a training set and a test set, and multiple machine learning algorithms are applied to develop predictive models of student success. The modeling phase is central to the research, where the application of machine learning algorithms – Random Forest (RF) and Gradient Boosted Trees (GBT) – determines the role and significance of stable features, as well as their contribution to predicting model performance.

RF is an ensemble of decision trees where each tree relies on randomly selected features from the input data. This approach reduces dependency among data points and ensures robustness by minimizing sensitivity to noise in the data, such as missing values and outliers [5], [9].

GBT is a machine learning algorithm that combines multiple weak predictive models (decision trees) to create a strong predictive model [10]. In GBT, trees are built iteratively, with each new tree attempting to correct and minimize the errors of the previous one [10].

Many consulted studies apply these algorithms to similar types of tasks, which is why they are applied in this research. Study [11] states that the GBT algorithm provides the best prediction for student success and failure, while [5] reports similar findings for the RF algorithm. Decision tree-based algorithms have also been applied in studies [12] and [13].

The results obtained using the algorithms are evaluated using metrics commonly used in regression tasks, namely Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Normalized Root Mean Squared Error (NRMSE). All metrics measure prediction errors, where lower values indicate better model performance. Based on the sensitivity analysis, insights into key features for prediction are obtained, and the resulting model will then be applied to predict new cases.

4 Results and Discussion

This section presents the results of the pilot study based on an analysis of data from the academic years 2023/2024 and 2024/2025.

Table 1. Presentation of results obtained using RF and GBT algorithms

Algorithm	Metric	2023/2024	2024/2025
Random Forest	MAE	2.3474	2.38
	RMSE	2.8221	2.8303
	NRMSE	0.2352	0.2881
Gradient Boosted Tree	MAE	2.2011	2.3093
	RMSE	2.7050	2.9587
	NRMSE	0.2254	0.3012

On the prepared dataset, after modelling with the mentioned algorithms, an evaluation is conducted and presented in Table 1.

Table 2. Presentation of the top 10 features obtained through sensitivity analysis, sorted by importance

RF - 2023/2024	RF - 2024/2025	GBT - 2023/2024.	GBT - 2024/2025
topinion: 0.3839	**topinion:** 0.3457	**topinion:** 0.4530	**topinion:** 0.3382
Exercise: 0.0961	**Assignment: Submission of the task for Week 3 exercises (Exercise 2 and 3) time:** 0.0480	**Excercise:** 0.1740	**Assignment: Submission of the task for Week 3 exercises (Exercise 2 and 3) time:** 0.0433
Assignment: Submission of the task for lab exercises - ERA model time: 0.0481	**activity_sum*activity_start_learning_before_exam_sum:** 0.0303	**Research seminar:** 0.0641	**Assignment: Submission of the task for lab exercises - Relational schema time:** 0.0402
Research seminar: 0.0340	**AssignmentMean:** 0.0297	**weekend_10:** 0.0300	**activity_start_learning_before_exam_Afternoon**: 0.0399
start_learning_before_exam: 0.0287	**activities_during_learning:** 0.0287	**start_learning_before_exam:** 0.0271	**AssignmentMean**: 0.0388
weekend_10: 0.0260	**Assignment: Submission of the task for lab exercises - Relational schema time:** 0.0260	**Assignment: Submission of the task for lab exercises - ERA model time:** 0.0269	**activity_start_learning_before_exam_sum:** 0.0370
Assignment: Submission of the task for lab exercises - Relational schema time: 0.0248	**Assignment: Submission of the task for lab exercises 4 - Project management time:** 0.0185	**median_periodicity_h:** 0.0250	**mode_periodicity_h:** 0.0358
activity_start_learning_before_exam_Afternoon: 0.0212	**Assignment: Oracle Apex submission 1 time:** 0.0184	**activity_Afternoon:** 0.0189	**Assignment: Submission of the task for lab exercises 4 - Project management time:** 0.0242
Assignment: Submission of the task for lab exercises 5 - BPMN time: 0.0174	**Assignment: Submission of the task for lab exercises 6 - Use case diagrams time:** 0.0179	**Assignment: Submission of the task for lab exercises - Relational schema time:** 0.0166	**activity_sum*activity_start_learning_before_exam_sum:** 0.0227
Assignment: Submission of the task for Week 3 exercises (Exercise 2 and 3) time: 0.0171	**start_learning_before_exam:** 0.0178	**active ty_start_learning_before_exam_Afternoon:** 0.0142	**Assignment: Submission of the task for lab exer-cises - ERA model time:** 0.0223

Table 2 presents the top 10 features identified as the most important during the sensitivity analysis, classified by years and algorithms.
So, one measure of stability, mentioned in study [9], is sensitivity analysis, which tracks features that consistently appear over time. Another measure is the coefficient of variation. Study [16] states that the coefficient of variation is a measure of dispersion used to detect changes in accuracy and monitor process stability. In this case, it is applied to assess feature stability over two years.

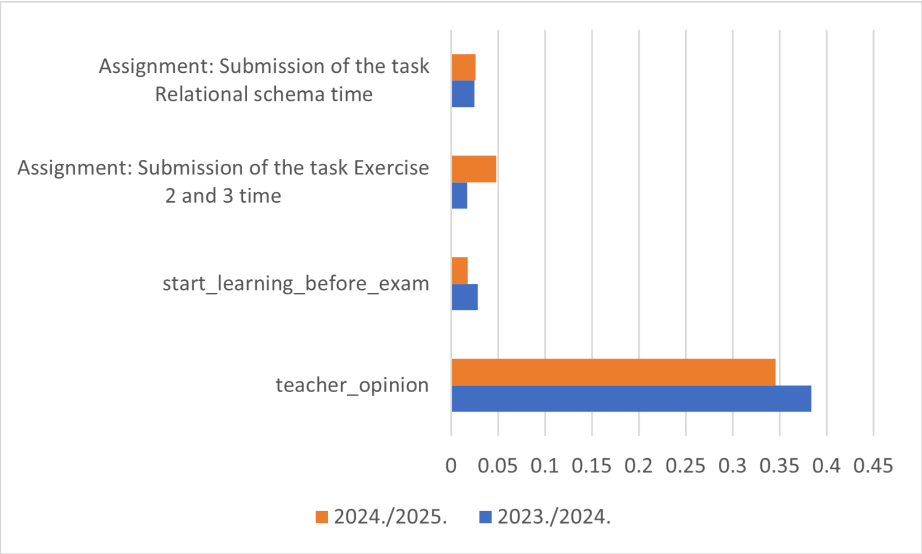

Fig. 2. Feature stability trends across years obtained using the RF algorithm

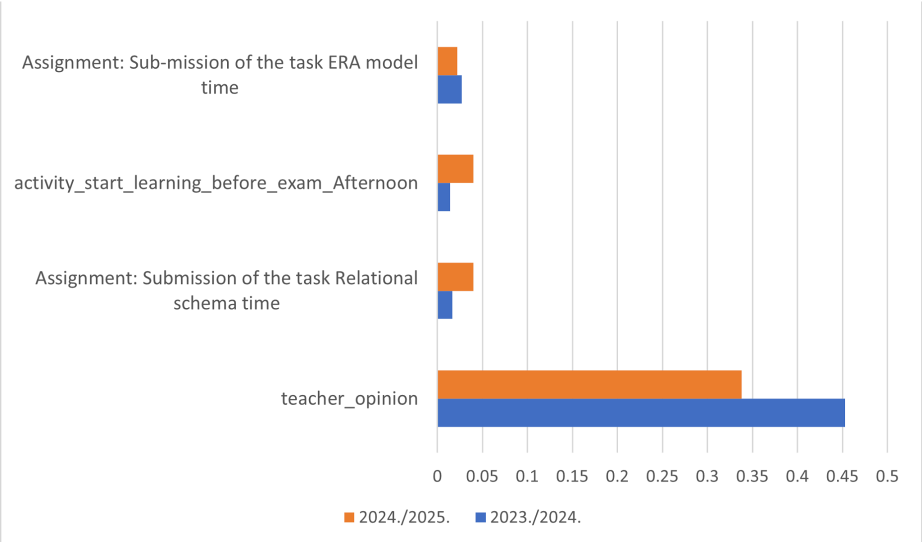

Fig. 3. Feature stability trends across years obtained using the GBT algorithm

The Figures 2 and 3 illustrate feature stability trends across years for four consistent features. The first image shows results obtained using the Random Forest (RF) algorithm, while the second reflects outcomes from Gradient Boosted Trees (GBT). The comparison highlights how feature importance and consistency vary between models, offering insights into their robustness and temporal reliability.

Table 3. Presentation of the coefficient of variation for features recurring in 2023/2024 and 2024/2025

Features in 2023/2024 and 2024/2025 - RF	Coefficient of variation	Features in 2023/2024 and 2024/2025 - GBT	Coefficient of variation
teacher_opinion	0.0524	teacher_opinion	0.1451
start_learning_before_exam	0.2344	Assignment: Submission of the task for lab exer-cises - Relational schema time	0.4155
Assignment: Submission of the task for Week 3 exercises (Exercise 2 and 3) time	0.4747	activity_start_learning_before_exam_Afternoon	0.4750
Assignment: Submission of the task for lab exercises - Relational schema time	0.0236	Assignment: Sub-mission of the task for lab exercises - ERA model time	0.0935

Table 3 presents the results of the pilot study for the coefficient of variation measure. The results show features that appear in both years for the RF and GBT algorithms, considering only the top 10 features identified through sensitivity analysis. Findings indicate that certain course assignments play a significant role in prediction, followed by the subjective teacher's opinion.

5 Conclusion

This study provides valuable insights into the temporal stability of student success predictors gained from multiple data sources. Reserach results highlight the

importance of identifying robust features to ensure the long-term reliability of predictive models in educational settings.
LMS features: *Assignment: Submission of the task for lab exercises - Relational schema time* (using the RF algorithm) and - *Assignment: Submission of the task for lab exercises - ERA model time* (using the GBT algorithm), which belong to the formative assessment type, are the most stable predictors of success, demonstrating the highest predictive power. This conclusion is derived from answering the research question posed:

The first question examines which group of integrated features contributes most significantly to model stability. The pilot study found that features related to teacher opinions and academic success have a greater impact on model stability.

The second question aims to identify which success predictors remain stable within an e-learning course over several years, measured by the coefficient of variation. The results show that within the e-learning system, assignments, learning start time, and activities in the system during learning can be extracted as stable predictors over time.

The third question aims to identify stable data features with strong predictive power in models generated using various machine learning algorithms, based on sensitivity analysis. Some of the features that can be identified include teacher opinions, assignments, activities during learning, and weekend activity patterns.

Additionally, the fourth research question examines features from the formative assessment group that have been identified as stable. These features belong to the academic success category. Among the extracted LMS features, assignments stand out, as sensitivity analysis indicates, they have higher predictive power. Furthermore, the coefficient of variation shows that some assignments have remained stable over the years.

It can be concluded that the study's findings on predictor significance align with the consulted research, with further contributions focusing on verifying the stability of features. i.e., success predictors.

The study's limitations are reflected in the small dataset. However, future research will expand to include data covering more years. Additionally, in the continuation of the study, more machine learning algorithms will be applied, and evaluations will be conducted across multiple metrics.

The future goal is also to collect data over multiple years within a single course before expanding the research to multiple courses.

Acknowledgement

This paper is supported by Croatian Science Foundation under the project SIMON: Intelligent system for automatic selection and machine learning algorithms in social sciences, UIP-2020-02-6312.

Declaration on Generative AI

During the preparation of this work, the author(s) used Copilot and Grammarly in order to: Edit references and grammar and spelling check. After using tool(s)/service(s), reviewed and edited the content as needed, and take(s) full responsibility for the publication's content.

References

1. Society for Learning Analytics (SoLAR): What is Learning Analytics? Available: https://www.solaresearch.org/about/what-is-learning-analytics/, last accessed 2025/07/01.
2. Conijn, R., Snijders, C., Kleingeld, A., Matzat, U.: Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. In: IEEE Transactions on Learning Technologies, vol. 10, no. 1, pp. 17–29 (2017).
3. Open-meteo: Free Weather API. Available: https://open-meteo.com/en/docs, last accessed 2025/03/21.
4. Saqr, M., Fors, U., Tedre, M.: How Learning Analytics Can Early Predict Under-Achieving Students in a Blended Medical Education Course. Medical Teacher 39(7), 757–767 (2017).
5. Saqr, M., López Pernas, S.: Why Explainable AI May Not Be Enough: Predictions and Mispredictions in Decision Making in Education. Smart Learning Environments 11(52), 1–16 (2024). Available: https://doi.org/10.1186/s40561-024-00343-4.
6. Abdulkareem Shafiq, D., Marjani, M., Ariyaluran Habeeb, R.A., Asirvatham, D.: Digital Footprints of Academic Success: An Empirical Analysis of Moodle Logs and Traditional Factors for Student Performance. Education Sciences 15(3), 304 (2025). Available: https://doi.org/10.3390/educsci15030304.
7. Klonkay, P.: Procjena indeksa kvalitete zraka za lebdeće čestice PM10 i PM2,5 grada Zagreba primjenom metoda strojnog učenja (Diplomski rad). Fakultet kemijskog inženjerstva i tehnologije (2024).
8. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 - Step-by-Step Data Mining Guide (2000). Available: https://mineracaodedados.files.wordpress.com/2012/12/crisp-dm-1-0.pdf.
9. Breiman, L.: Random Forests. Machine Learning 45, 5–32 (2001).
10. Shah, J.: Gradient Boosting. Technical Report, no. 1–6, Computer Science Department, Nirma University, Ahmedabad, India (2020).
11. Blessed Deho, O., Joksimovic, S., et al.: Should Learning Analytics Models Include Sensitive Attributes? Explaining the Why. IEEE Transactions on Learning Technologies 16(4), 560–571 (2023).
12. Balabaeva, K., Kovalchuk, S.: Comparison of Temporal and Non-Temporal Features Effect on Machine Learning Models Quality and Interpretability for Chronic Heart Failure Patients. Procedia Computer Science 156, 87–96 (2019).
13. Teinemaa, I., Dumas, M., et al.: Temporal Stability in Predictive Process Monitoring. Data Mining and Knowledge Discovery 32, 1306–1338 (2018). Available: https://doi.org/10.1007/s10618-018-0575-9.
14. Chai, T., Draxler, R.R.: Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? – Arguments Against Avoiding RMSE in the Literature. Geoscientific Model Development 7, 1247–1250 (2014).
15. Lifesight: What is Normalized Root Mean Square Error (NRMSE)?. Available: https://lifesight.io/glossary/normalized-root-mean-square-error/, last accessed 2025/04/07.
16. Jalilibal, Z., Amiri, A., et al.: Monitoring the Coefficient of Variation: A Literature Review. Computers & Industrial Engineering 161, 107600 (2021).
17. Van Rossum, G., Drake, F.L.: Python 3 Reference Manual. Scotts Valley, CA: CreateSpace (2009).