Student Success Prediction Based on Machine Learning and Learning Styles

Dijana Oreški
University of Zagreb,
Faculty of Organization and Informatics,
Pavlinska 2, Varazdin,
Croatia

Received: May 27, 2024. Revised: March 11, 2025. Accepted: May 11, 2025. Published: July 18, 2025.

Abstract—This study aims to develop a predictive model for student success by integrating machine learning algorithms with learning style analysis. Educational institutions increasingly recognize the value of early performance prediction to implement timely interventions and enhance learning outcomes. Learning management systems generate vast amounts of data. The proposed research will analyze student interaction data from Moodle learning management system, including course logins, resource access patterns, assignment submissions, and assessment performance. These digital footprints will be combined with learning style assessments to identify patterns of academic achievement. Machine learning algorithms are applied and compared to determine the most effective predictive model. This research contributes to educational data mining by exploring the intersection between digital behavior patterns, individual learning preferences, and academic outcomes. The resulting model achieves high prediction accuracy, enabling proactive educational interventions that adapt to students' learning styles while leveraging Moodle's AI capabilities for personalized learning experiences.

Keywords—Machine learning, educational data mining, student success, predictive modelling, decision tree, sensitivity analysis.

I. Introduction

THIS research is performed under project SIMON: Intelligent system for automatic selection of machine learning algorithms in social sciences. One of the projects main aims is investigation of machine learning algorithms performance in education. In recent years, the application of machine learning techniques in education has gained significant attention, especially for the student success prediction. This growing interest comes from the potential to enhance educational outcomes by identifying at-risk students early and providing focused interventions. The intersection of machine learning and learning management system data presents a promising research path for developing more accurate and personalized student success prediction models. By incorporating data on students' preferred learning modalities alongside traditional academic indicators, these models aim to provide a deeper view of student potential and challenges.

This research explores the integration of machine learning algorithms with learning management system Moodle data and student learning style preferences to predict students' results and development of a comprehensive approach for predicting student success. The study aims to address several key questions:

i. How can machine learning models leverage effectively learning management system data and students learning styles data to develop predictive models with improved prediction accuracy?

ii. What are the most significant predictors of student success when considering learning management system data?

iii. How can the insights gained from these predictive models be translated into actionable strategies for educators and institutions?

By examining these questions, this study seeks to contribute to the field of educational data mining and learning analytics, while also providing practical implications for enhancing student support systems and instructional design.

This paper is organized as follows. Section 2 gives related work. Section 3 provides data and methodological insights. Section 4 shows research results, whereas section 5 discusses implications of results for science and profession taking into account limitations of the paper. Section 6 concludes the paper.

II. Related Work

Literature review incorporated following aspects:

• Educational domain: examining study papers focusing on higher education students using Moodle as the learning management system?

• Machine Learning Application: examining study papers employing machine learning techniques specifically for predicting student performance?

• Outcome Measurement: examining study papers which measure and report academic performance or student success outcomes?

• Data Sources: examining study papers that analyze actual student interaction data collected from Moodle?

Several studies highlight the importance of specific Moodle features in predicting student success: e.g. [1] found that quiz-related predictors have the largest impact on predicting student success. This aligns with findings from other studies that emphasize the importance of student engagement metrics, such as resource views and quiz attempts. Authors in [2] provide a more detailed analysis, showing that different features may be more or less important at different stages of a course. This suggests that the predictive power of certain features may change over time. The timing of predictions varies across studies, with some focusing on early prediction and others making predictions throughout the course. Authors in [3] that it's possible to detect at-risk students as early as the first week of a course, with Area under the Curve of the Receiver Operating Characteristic (AUC ROC) values ranging from 0.7 to 0.9.

[1] provide a more detailed view, comparing predictions at 25%, 50%, 75%, and 100% of course completion. Their findings indicate that prediction accuracy can vary at different stages, with different algorithms performing best at different points in the course.

The studies utilize a range of Moodle data for their predictions, including: student activities and engagement metrics, test scores and quiz performance, resource access (e.g., lecture notes, videos), and communication data (e.g., emails, forum participation). The variety of data types used suggests that a comprehensive approach to data collection within Moodle and integration with other sources may be beneficial for prediction accuracy.

Several factors emerge as important considerations for model selection:

• Prediction timing: Some models perform better for early prediction, while others are more accurate later in the course.

• Feature availability: The choice of model may depend on which Moodle features are available or considered most relevant for a particular context.

• Interpretability: While not explicitly discussed in most abstracts, the interpretability of models (e.g., decision trees vs. neural networks) may be an important factor for practical implementation.

• Computational resources: The complexity of models like neural networks may require more computational resources compared to simpler algorithms.

• Course context: The effectiveness of different models may vary depending on the specific course or discipline context.

These factors suggest that a one-size-fits-all approach to student success prediction may not be optimal, and that model selection should be adapted to the specific needs and constraints of each educational context.

III. Research Methods

This paper employs CRISP DM standard methodology consisting of domain understanding, data understanding, data preparation, model development, performance evaluation and deployment. Domain understanding is presented through literature review in previous section. Deployment is implemented through research discussion. CRISP DM activities of data understanding, data preparation, modelling and evaluation are presented in this section.

A. Data understanding

In this study, we decision tree algorithm is implemented with an aim to predict student success based on learning styles and activity data collected from the Moodle Learning Management System (LMS). Decision trees approach has several advantages over other machine learning approaches. Some of the advantages are: their interpretability, ability to handle both categorical and numerical data, and ability to model non-linear relationships. The data for this study was extracted from the Moodle LMS, which captures a comprehensive range of student interactions with the learning platform. The dataset includes:

(i) Student activity logs (frequency and duration of logins, resource access patterns).

(ii) Engagement with different types of learning materials (videos, readings, interactive exercises).

(iii) Completion rates of assignments and assessments.

(iv) Participation in discussion forums and collaborative activities.

(v) Timestamps of activities to analyze temporal patterns.

Prior modelling, preprocessing steps were performed.

B. Decision tree modelling

The decision tree algorithm was used, more specifically, C4.5 algorithm. The C4.5 algorithm was implemented because this approach can handle both continuous and categorical attributes and has effective pruning mechanisms. The dataset was split into training (70%) and testing (30%) sets to evaluate model performance. Grid search with cross-validation was employed to optimize key parameters:

- Maximum tree depth.
- Minimum samples required for node splitting.
- Minimum samples required for leaf nodes
- Split quality criterion (Gini impurity vs. information gain).

DOI: 10.46300/9109.2025.19.10 Volume 19, 2025

Post-pruning methods were applied to prevent overfitting and improve generalization. 10-fold cross-validation was used to ensure robust performance evaluation.

C. Model evaluation

The model's predictive performance was assessed using accuracy. A key innovation in this methodology is the incorporation of learning style data as features in the decision tree model. This allows for exploration of how different learning preferences interact with online learning behaviors to influence academic outcomes. The learning style dimensions were analyzed both as independent predictors and in combination with behavioral LMS data to identify potential interaction effects.

IV. Research results and discussion

Predictive models of student success achieved accuracy of 94.6%. Beyond predictive performance, significant attention was devoted to model interpretability. The resulting decision trees were analyzed to extract actionable insights about: (i) the most influential factors in predicting student success, critical thresholds for intervention, (iii) interaction patterns between learning styles and online behaviors.

This methodology enables not only accurate prediction of student outcomes but also provides transparent, explainable results that can inform educational interventions and personalized learning approaches.

The integration of different data sources is important for effective student success prediction models. Educational institutions can leverage data from multiple sources, including Moodle learning management systems, which store student interaction data such as course logins, resource access patterns, assignment submissions, and assessment performance, [4]. These digital footprints provide valuable insights when combined with other student data to identify patterns between preferred learning modalities and academic achievement. Data preparation takes away majority of the time of the data mining process, between 65% and 90%, [5]. The process involves data integration and data transformation, a foundational requirement for machine learning implementation, [6]. There are several challenges for institutions regarding dispersion of data between different data sources. Additionally, maintaining data integrity in the modelling process is crucial for accurate and reliable predictions. For blended learning environments that combine offline and online classrooms, researchers can obtain more comprehensive attributions for student performance from multiple data sources, resulting in better prediction performance. This multi-source approach has been shown to improve prediction accuracy compared to single data source models.

Several machine learning algorithms have demonstrated effectiveness in predicting student performance, [7]. A decision tree is among most used machine learning approaches based on their explain ability and high prediction performance.

Other commonly used algorithms include Artificial Neural Networks (ANN), Naïve Bayes, and SVM, [8]. Research indicates that ensemble methods combining multiple algorithms often outperform one algorithm approaches by 10-15% in educational prediction tasks. However, the Stacking Regressor, while effective, did not significantly outperform its base models in one study indicating that complex models do not necessarily lead to better predictive performance.

Successful deployment of machine learning models mostly depends on how well predictive models are integrate with existing educational systems, [9]. Best practices include developing implementation roadmaps, establishing transparent data governance policies, and creating user-friendly interfaces for stakeholders. Deployment challenges extend beyond technical considerations to include organizational and procedural alignment. Higher education institutions must address several obstacles:

Infrastructure challenges: Solutions are often highly individual and require substantial technical knowledge and skills. The lack of standardized machine learning case studies requires development of individual infrastructure and approach for many projects, significantly complicating deployment.

Ongoing validation: Deployed models require continuous validation. Currently, validation is often performed through manual approach requiring the combined knowledge of data scientists and domain experts, making it time-consuming and very complicated.

Data drift: Changes in data distribution over time affect model performance significantly. While some models can automatically detect data drifts, they are usually handled through manual model adjustments rather than automated updates.

Legal and ethical considerations: Legal requirements often pose significant challenges, particularly regarding data privacy protection and decisions about who is responsible for algorithmic decisions.

Effective training programs are essential for successful implementation, combining technical knowledge with pedagogical context. For sustainable implementation, institutions should establish regular model review cycles to assess whether current models still address institutional needs. Predictive models usually require significant recalibration every few years to maintain optimal performance.

Evidence shows that when properly implemented, predictive modelling can dramatically improve educational outcomes, [10]. Studies reveal that institutions using predictive analytics improved their student retention rates.

V. Conclusion

Predicting student success using machine learning (ML) and learning style analysis combines behavioral patterns, academic data, and pedagogical context to create actionable insights. While machine learning offers powerful tools for predicting student success, effective implementation requires balancing model precision with pedagogical context. Learning styles add interpretability but must be contextualized within broader behavioral and academic datasets. Future systems prioritizing explainability, adaptability, and ethical design will likely yield the most impactful educational interventions. For the future research directions following aspects should be considered:

(i) explainable AI (XAI): Developing hybrid architectures that explain why specific learning styles or behaviors correlate with success.

(ii) real-time adaptation: Integrating predictive models into adaptive learning systems to dynamically adjust content delivery based on style-performance connection.

(iii) Cross-institutional benchmarks: Standardizing datasets and evaluation metrics to improve model comparability and robustness.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

The author wrote, reviewed and edited the content as needed and verifies that none utilised artificial intelligence (AI) tools were used. The author takes full responsibility for the content of the publication.

References

[1] L. Bognár and T. Fauszt, “Different learning predictors and their effects for Moodle machine learning models,” in *Proc. 11th IEEE Int. Conf. Cognitive Infocommunications (CogInfoCom)*, Mariehamn, Finland, Sept. 23–25, 2020, pp. 405–410. doi:10.1109/CogInfoCom50708.2020.9284432

[2] C. Kaensar and W. Wongnin, “Analysis and prediction of student performance based on Moodle log data using machine learning techniques,” *Int. J. Emerg. Technol. Learn. (iJET)*, vol. 18, no. 10, pp. –, 2023. doi:10.3991/ijet.v18i10.35841

[3] L. A. Buschetto Macarini, C. Cechinel, M. F. Batista Machado, V. F. Culmant Ramos, and R. Muñoz, “Predicting students' success in blended learning—Evaluating different interactions inside learning management systems,” *Appl. Sci.*, vol. 9, no. 15, p. 3138, 2019.

[4] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, “Educational data mining techniques for student performance prediction: Method review and comparison analysis,” *Front. Psychol.*, vol. 12, Art. no. 698490, 2021.

[5] L. Baier, F. Jöhren, and S. Seebacher, “Challenges in the deployment and operation of machine learning in practice,” in *Proc. 27th Eur. Conf. Inf. Syst. (ECIS)*, Stockholm, Sweden, Jun. 2019, vol. 1.

[6] L. U. Xi, “Modern Education: Advanced Prediction Techniques for Student Achievement Data,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 1, pp. –, 2024. doi:10.14569/ijacsa.2024.01501126

[7] S. Gunasekara and M. Saarela, “Explainability in educational data mining and learning analytics: An umbrella review,” in *Proc. 17th Int. Conf. Educ. Data Mining (EDM)*, 2024, pp. 887–892.

[8] N. R. Yadav and S. S. Deshmukh, “Prediction of student performance using machine learning techniques: A review,” in *Proc. Int. Conf. Appl. Mach. Intell. Data Anal. (ICAMIDA)*, May 2023, pp. 735–741. Atlantis Press.

[9] D. Oreški, “Application of machine learning methods for data analytics in social sciences,” *WSEAS Trans. Syst.*, vol. 22, pp. 69–72, Mar. 7, 2023. doi: 10.37394/23202.2023.22.8

[10] R. Kovač and D. Oreški, “Educational data driven decision making: Early identification of students at risk by means of machine learning,” in *Proc. 29th Central European Conf. Inf. Intell. Syst. (CECIIS)*, Varaždin, Croatia, Sept. 19–21, 2018, pp. 231–237.

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

Author Contributions: Dijana Oreski organized research, carried out experiments and wrote the paper.

Sources of funding for research presented in a scientific article or scientific article itself

This paper is supported by the Croatian Science Foundation under the project SIMON: Intelligent System for automatic selection of machine learning algorithms in Social Sciences, UIP-2020-02-6312.

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.