

Explainable Predictive Models in Learning Analytics: A Proposal for Identifying Performance Patterns at UNIVESP

Rafael Amorim Faleiros, José Alvelino Placca

Universidade Virtual do Estado de São Paulo (UNIVESP), São Paulo, Brasil

24200095@aluno.univesp.br, jose.placca@univesp.br

Abstract. *This paper proposes an explainable Learning Analytics (LA) model to identify performance patterns and dropout risks at UNIVESP. It integrates Virtual Learning Environment (VLE) logs and social media textual data, processing them via Explainable Artificial Intelligence (XAI). Aiming at an institution with over 90,000 students, the model incorporates social variables in strict adherence to data privacy (LGPD). By using Gradient Boosting and SHAP values, the approach overcomes the accuracy-interpretability trade-off, mitigating algorithmic aversion. This ensures predictions remain pedagogically actionable, fostering personalized interventions and transforming distance education management on a large scale.*

Keywords— *Learning Analytics, Explainable Artificial Intelligence (XAI), Predictive Models, Inclusive distance education, Student Engagement.*

Resumo. *Este artigo propõe um modelo explicável de Learning Analytics (LA) para identificar padrões de desempenho e riscos de evasão na UNIVESP. A abordagem integra logs do Ambiente Virtual de Aprendizagem (AVA) e dados textuais de redes sociais, processados via Inteligência Artificial Explicável (XAI). Para uma instituição com mais de 90.000 alunos, o modelo incorpora variáveis sociais em estrita conformidade com a LGPD. Ao utilizar Gradient Boosting e valores SHAP, supera-se o trade-off entre precisão e interpretabilidade, mitigando a aversão algorítmica. Isso garante previsões pedagogicamente acionáveis, fomentando intervenções personalizadas na gestão da educação a distância em larga escala.*

Palavras-chave— *Aprendizado Analítico, Inteligência Artificial Explicável (XAI), Modelos Preditivos, Educação a distância inclusiva, Engajamento Estudantil.*

1. Introduction

The expansion of Distance Learning (DL) in Brazil has established itself as a fundamental vector for the democratization of higher education, allowing historically marginalized population strata to access academic training. According to Moran [Moran 2015], this modality promotes inclusion as long as it is mediated by dialogic and interactive principles. However, as Mantoan [Mantoan 2003] warns, effective inclusion requires the continuous removal of structural, pedagogical, and technological barriers, which becomes particularly challenging in large-scale contexts. In this scenario, dropout emerges as a severe problem in DL, compromising individual trajectories, educational efficiency, and institutional sustainability. Studies such as [Tinto 1993] and [Rovai 2003] highlight that dropout rates in distance courses tend to be significantly higher

than those in on-site education, evidencing the multifactorial complexity of this phenomenon.

The problem of dropout in DL must be understood from a systemic perspective, involving both structural and individual factors. According to [Bean and Metzner 1985], it is possible to distinguish non-avoidable causes—such as changes in socioeconomic status, professional demands, or family issues—and avoidable causes, which can be mitigated through institutional interventions. Among the latter, pedagogical factors stand out especially, such as the low quality of instructional design, the absence of formative feedback, the limited interaction between students and teachers, and the lack of engagement promoted by didactic activities [Moore 2013, Garrison et al. 2000]. Such elements are central to student persistence, especially in virtual environments, where student autonomy requires continuous and intentional pedagogical support. Thus, dropout should not be treated merely as an inevitable side effect, but as a partially controllable phenomenon, whose reduction depends on evidence-based pedagogical strategies.

Currently, educational managers still face the strategic "trade-off" between the precision of Artificial Intelligence (AI) models and the human capacity to interpret such results for pedagogical purposes. The central problem lies in the opacity of traditional predictive models, which frequently fail to explain why a student is classified at risk of dropout, limiting their practical applicability. In this context, the main objective of this research is to develop a proposal for a predictive model that integrates structured performance data, VLE logs, and textual data from social networks. To support this proposal and overcome the opacity of traditional models, the study is guided by three central hypotheses: (H1) the correlation between the quality of interactions in the VLE and academic performance is enhanced by active and collaborative methodologies; (H2) negative sentiments expressed on social networks act as leading indicators of inactivity in the system; and (H3) the alignment between professional occupation and the area of study acts as a strong moderating factor of persistence. The Learning Analytics analysis presented here maintains that the identification of patterns on a large scale is only viable and ethically defensible when accompanied by algorithmic transparency, allowing raw data to be transformed into actionable pedagogical intelligence.

The remainder of this paper is organized as follows: Section 2 presents the theoretical framework, addressing the intersections between Learning Analytics, Explainable Artificial Intelligence (XAI), and dropout in DL. Section 3 details the UNIVESP data ecosystem and the categorization of the selected variables. Section 4 describes the proposed methodology and the architecture of the multimodal processing model. Section 5 outlines the research hypotheses and discusses the initial results of the approach. Finally, Section 6 consolidates the final considerations of the study.

2. Theoretical Framework

Learning Analytics (LA) has established itself as a strategic predictive layer in DL, allowing the anticipation of patterns of academic success or failure [Siemens and Baker 2012]. However, the adoption of complex predictive models [Baker and Inventado 2016, Romero and Ventura 2020] runs into the trade-off with interpretability, generating "algorithmic aversion" on the part of educators in the face of opaque systems [Dietvorst et al. 2015, Pischetola and Riedner 2021]. To overcome this barrier, Explainable

Artificial Intelligence (XAI) emerges as a fundamental link, making predictions interpretable [Adadi and Berrada 2018] and allowing the secure integration of multimodal data into effective pedagogical actions.

One of the main applications of these targeted pedagogical actions lies in combating dropout in DL. In this scenario, the literature distinguishes non-avoidable causes from avoidable pedagogical causes, such as the absence of interaction and unengaging instructional design [Bean and Metzner 1985, Moore 2013, Rovai 2003]. The mitigation of these factors demands the intentional use of active methodologies, such as Problem-Based Learning (PBL), and the strengthening of learning communities (Community of Inquiry), which promote social presence and student engagement [Garrison et al. 2000, Hrastinski 2009, Deterding et al. 2011]. However, while LA traditionally focuses on the behavioral trail restricted to the VLE to anticipate academic performance [Baker and Inventado 2016, Romero and Ventura 2020, Siemens and Baker 2012], it frequently fails to capture the student's emotional state that precedes the decline of this technical engagement.

To fill this gap, Sentiment Analysis (SA) has been applied to capture external negative socio-emotional states (such as anger, confusion, and boredom) that frequently precede inactivity [Bobó et al. 2019]. Since isolated sentiments do not show a strong direct linear correlation with dropout, this work proposes the structural convergence between LA and SA. By crossing indicators of frustration, extracted via NLP on social networks, with VLE performance data, the multimodal architecture transcends isolated analysis. Supported by XAI, this model transforms emotion into a robust indicator, ensuring that pedagogical interventions are accurately targeted at the root causes of disengagement.

3. The UNIVESP Ecosystem: Scale of Available Data

UNIVESP acts as a data laboratory of continental scale, with 90,000 students in 382 municipalities, reaching 93% of the state population. Its institutional maturity is evidenced by Working Groups (WGs) and the participation of deaf reviewers who validate the accessibility of content.

The selection of the model's variables is based on studies of Learning Analytics applied to dropout and academic performance [Baker and Inventado 2016, Romero and Ventura 2020]. The inclusion of social network data is justified to go beyond VLE data, by capturing the externalization of socio-emotional frustrations in real-time before they materialize as inactivity on the platform. The variables proposed for the model, integrating performance and assistive accessibility metrics, are detailed below:

Table 1. Proposed model variables.

Data Categories	Specific Variables	Analytical Impact
Cognitive Performance	Grades of activities and on-site exams.	Direct indicator of competence mastery.
Behavior in VLE	Frequency, session time, and navigation trails.	Measures persistence and study pace.
Digital Accessibility	Usage logs of NVDA, MathPlayer, and access to OERs.	Identifies technical barriers and effectiveness of assistive tools.

Social Interaction (NLP)	Georeferenced posts in social networks filtered by topics.	Acts as an early indicator (Leading Indicator) of demotivation.
Hub Infrastructure	Location, physical resources, and on-site technical support.	Mediating variable for engagement and access quality.
Professional Context	Current employment relationship (identified via NLP on professional networks) and the course area.	Moderating variable of persistence, measuring the practical value and motivation of the student.

To ensure analytical rigor and compliance with the LGPD, the correlation between sentiments captured on social networks and VLE data does not occur individually. The proposed architecture uses georeferenced groupings and analysis of post themes. Extracted posts are filtered by the geographic coordinates of each UNIVESP Hub, and then their emotional topics are crossed with the performance averages (grades and accesses) of the group of students enrolled in that specific hub. Although the implementation of this full multimodal pipeline is scope for future improvements, correlation based on group similarity allows the model to identify regional dropout risk trends without exposing students' identities.

4. Proposed Methodology and Model Architecture

From a methodological point of view, this research is characterized as applied and exploratory, with a qualitative-quantitative approach, aiming to propose and structure a multimodal predictive architecture for DL. The robustness of the proposal lies in a workflow divided into four technical stages and a governance layer:

4.1. Collection and Integration

Unification of VLE logs with the extraction and transformation of data from public social networks (via API/Scraping) through automated routines to ensure long-term data flow, focusing on topics linked to the institution through keyword searches

4.2. Multimodal Processing

Application of Natural Language Processing (NLP), using Sentiment Analysis and Latent Dirichlet Allocation (LDA) techniques to identify and separate topics of engagement or frustration, reducing sarcasm interpretation errors.

4.3. Data Engineering and Temporal Synchronization

To ensure multimodal integration despite the asynchrony in data sources, the architecture proposes a synchronization based on a temporal window. The raw data extracted from social networks are grouped into weekly packages, containing the publication's timestamp, the nearest Hub (obtained by relating coordinates), and the sentiment score generated by the NLP. In the structured database, these weekly packages act as independent variables that are overlaid onto the timeline of VLE accesses and grades for that same week and region. This time discretization aligns the sampling frequencies, allowing the algorithm to identify if a peak of negative sentiment in a given week reflects a drop in VLE attendance in the following week, thus enabling early prediction.

4.4. Modeling and Explainability

Implementation of Gradient Boosting or Decision Tree algorithms, using SHAP (SHapley Additive exPlanations) values to quantify the contribution of each variable (e.g., the relevance of manifestations of discontent in social networks (identified via NLP in public texts) together with reduced attendance in the VLE for defining the danger of school abandonment).

4.5. Qualitative Validation and Feedback Loop

After the generation of explainable outputs, the results are confronted with feedback from the WGs and deaf reviewers at UNIVESP based on institutional accessibility protocols [Oliveira et al. 2025]. This process ensures that the pedagogical relevance of the predictions is validated by specialists, guaranteeing that the actions suggested by the model translate into real interventions that promote educational equity.

Figure 1 summarizes the complete architecture and the multimodal processing flow described above.

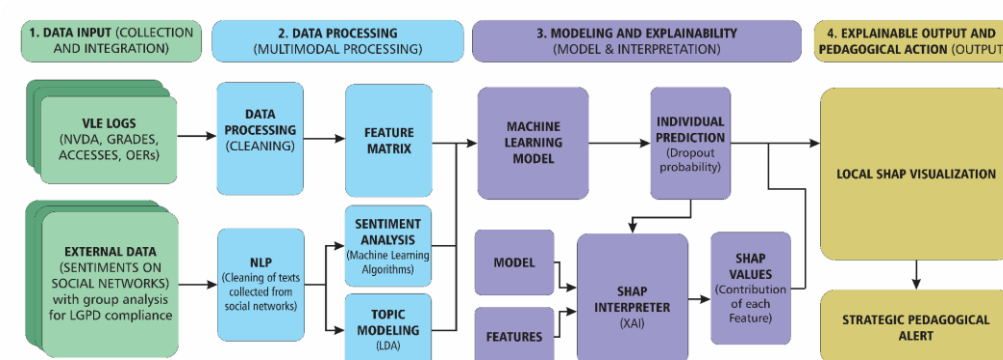


Figure 1. Multimodal Processing Flow and SHAP Explainability. Source: Own elaboration (2026).

The research foresees the mandatory anonymization of social data and free and informed consent. As [Alegre and Pinheiro 2019] warn, data collection on digital platforms requires an ethical protocol that respects the boundary between the public and the private, guaranteeing anonymity and non-authorship of the processed posts. The focus is on the analysis of group trends to avoid individual stigmatization, ensuring that digital monitoring fully respects student privacy and LGPD guidelines.

5. Expected Results and Initial Discussion

The project aims to validate hypotheses that relate interaction in the virtual environment, social context, and academic performance:

5.1. Hypothesis 1

The correlation between the volume and quality of interactions in the Virtual Learning Environment (VLE) and academic performance is mediated by the presence of interactive didactic materials and the use of educational technologies that favor active methodologies, especially Problem-Based Learning (PBL) activities performed in

groups. Environments that promote collaboration, interactivity, and collective knowledge construction tend to enhance student engagement, increasing the effectiveness of study time and contributing positively to academic performance.

5.2. Hypothesis 2

Negative sentiment or isolation expressed on social networks acts as a leading indicator, preceding inactivity in the Virtual Learning Environment (VLE) in a time interval that allows for preventive interventions before the first low grade, as corroborated by recent research in AI and educational warning systems [Bobó et al. 2019, Rout et al. 2025].

5.3. Hypothesis 3

Employability status and the job's adherence to the training area (identified via NLP on professional social networks) act as strong predictors of persistence. The crossing of this educational data in groups works as a moderating variable that reinforces engagement and the perception of the practical value of the degree, mitigating dropout risks associated with external professional demands.

The initial discussion suggests that Explainable Artificial Intelligence (XAI) will allow UNIVESP to implement more contextualized strategic alerts, such as: "Student with high risk associated with low levels of interaction in collaborative activities and signs of frustration in external networks". This contributes to the transformation of institutional culture, moving it from a reactive and punitive approach to a proactive, pedagogical, and humanized monitoring model centered on engagement and strengthening the student's sense of belonging.

6. Final Considerations

The integration of explainable predictive models aligns directly with UNIVESP's mission to promote technological innovation with social commitment and educational inclusion, as discussed by [Moreira and Schlemmer 2020]. By articulating Learning Analytics, predictive models, and Explainable Artificial Intelligence (XAI), the proposal moves towards an "onlife digital education," in which the boundaries between the physical and virtual become increasingly blurred, without losing sight of the complexity of the subjects who learn. In this sense, the student ceases to be just a set of data and comes to be understood in their cognitive, social, and contextual dimension, especially when considering variables such as interaction in the VLE, engagement in collaborative activities, and non-identifiable social conditions.

The expected results suggest that the combination of predictive capacity and interpretability can significantly transform pedagogical management in DL. By identifying early patterns associated with dropout risk — especially those linked to avoidable pedagogical causes such as low engagement, absence of meaningful interaction, and weaknesses in instructional design — the institution gains concrete subsidies for more effective interventions. The role of active methodologies, such as Problem-Based Learning (PBL), and interactive educational technologies, which favor collaborative learning and strengthen the student's sense of belonging, is highlighted in this context. However, the effectiveness of these strategies is intrinsically linked to the student's digital literacy, a factor that must be monitored and promoted institutionally to

ensure that technology acts as a facilitator of persistence, and not as an element of technical exclusion.

Additionally, by incorporating social network data through the analysis of similar profiles — preserving individual privacy [Alegre and Pinheiro 2019] — the model expands its capacity to capture socio-emotional signals that often precede academic disengagement. Explainability, in turn, ensures that these insights are not only technically robust but also pedagogically actionable, reducing algorithmic aversion and promoting greater trust on the part of tutors and managers.

However, the proposal faces limitations. Dependence on external data is subject to technical and institutional restrictions, such as changes in API access policies. Furthermore, the effectiveness of AI-based interventions directly depends on the continuous training of teachers and their ability to interpret and apply the results in everyday pedagogical practice. At this point, the importance of institutional initiatives aimed at developing digital and analytical skills is reinforced, as recommended in recent UNIVESP guidelines.

As a consequence, the project reinforces the need for an institutional data-oriented culture, where pedagogical decisions are based on evidence and strict ethical principles. Compliance with the General Data Protection Law (LGPD) is not only a legal requirement but a structuring element of trust in the use of educational technologies.

In the final analysis, Explainable Artificial Intelligence is a strategic tool for promoting educational equity. By making visible the factors that influence performance and dropout, technology ceases to be an opaque instrument and begins to act as a mediator of more inclusive pedagogical practices, sensitive to differences and oriented toward student success. Thus, the project contributes to the advancement of São Paulo's public education not only in scale but in quality, ensuring that technological innovation expands — rather than invisibilizes — the multiple realities that compose its educational ecosystem.

References

- Adadi, A. and Berrada, M. (2018) “Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)”, *IEEE Access*, v. 6, p. 52138-52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Alegre, P. and Pinheiro, P. (2019) “Uma análise sobre a utilização de dados coletados nas redes sociais online para a pesquisa acadêmica”, *Revista Ibérica de Sistemas e Tecnologias de Informação*.
- Baker, R. S. and Inventado, P. S. (2016) “Educational Data Mining and Learning Analytics”, In: *Emergence and Innovation in Digital Learning*, Athabasca University Press, p. 83-98. https://doi.org/10.1007/978-1-4614-3305-7_4.
- Bean, J. P. and Metzner, B. S. (1985) “A Conceptual Model of Nontraditional Undergraduate Student Attrition”, *Review of Educational Research*, v. 55, n. 4, p. 485-540. <https://doi.org/10.2307/1170245>.
- Bobó, M., Campos, F., Ströele, V., Braga, R. and David, J. M. N. (2019) “Análise de Sentimento na Educação: Um Mapeamento Sistemático da Literatura”, In: *Simpósio*

XXI Latin-American Conference on Learning Technologies (LACLO 2026)

- Brasileiro de Informática na Educação (SBIE), SBC, p. 249-258. <https://doi.org/10.5753/cbie.sbie.2019.249>.
- Deterding, S. et al. (2011) “From game design elements to gamefulness: defining gamification”, In: Proceedings of the 15th MindTrek Conference, ACM, p. 9-15. <https://doi.org/10.1145/2181037.2181040>.
- Dietvorst, B. J., Simmons, J. P. and Massey, C. (2015) “Algorithm aversion: People erroneously avoid algorithms after seeing them err”, *Journal of Experimental Psychology: General*, v. 144, n. 1, p. 114-126. <https://doi.org/10.2139/ssrn.2466040>.
- Garrison, D. R., Anderson, T. and Archer, W. (2000) “Critical inquiry in a text-based environment: Computer conferencing in higher education”, *The Internet and Higher Education*, v. 2, n. 2-3, p. 87-105. [https://doi.org/10.1016/s1096-7516\(00\)00016-6](https://doi.org/10.1016/s1096-7516(00)00016-6).
- Hrastinski, S. (2009) “A theory of online learning as online presence”, *Computers & Education*, v. 52, n. 1, p. 78-82.
- Mantoan, M. T. E. (2003) “Inclusão escolar: o que é? por quê? como fazer?”, São Paulo: Moderna.
- Moore, M. G. (2013) “The theory of transactional distance”, In: *Handbook of distance education*, Edited by M. G. Moore, 3. ed., New York: Routledge, p. 66-85. <https://doi.org/10.4324/9780203803738.ch5>.
- Moran, J. M. (2015) “Mudando a educação com metodologias ativas”, In: *Convergências Midiáticas, Educação e Cidadania*, Edited by C. A. Souza and S. T. Morales, Ponta Grossa: FUEPG.
- Moreira, J. A. and Schlemmer, E. (2020) “Por um novo conceito e paradigma de educação digital onlife”, *Revista UFG, Goiânia*, v. 20, n. 26, p. 1-20. <https://doi.org/10.5216/revufg.v20.63438>.
- Oliveira, E. A., Placca, J. A. and Conti, T. C. B. (2025) “Acessibilidade no ensino a distância: experiências e práticas inclusivas na UNIVESP”, *ESUD/CIESUD*. <https://submissoes.site.ufabc.edu.br/index.php/esud2025/article/view/147>
- Pischetola, M. and Riedner, D. D. T. (2021) “A inovação das práticas pedagógicas com uso de tecnologias digitais no ensino superior”, *ETD - Educação Temática Digital*, v. 23, n. 1, p. 210-228. <https://doi.org/10.20396/etd.v23i1.8655732>.
- Romero, C. and Ventura, S. (2020) “Educational data mining and learning analytics: An updated survey”, *WIRES Data Mining and Knowledge Discovery*, v. 10, n. 3, e1355. <https://doi.org/10.1002/widm.1355>.
- Rout, D. et al. (2025) “Inspecting the Paradigm of Artificial Intelligence in Sentiment Analysis of Learners’ Pursuance in Higher Education: A Review”, p. 326-343. https://doi.org/10.2991/978-94-6463-950-6_22.
- Rovai, A. P. (2003) “In search of higher persistence rates in distance education online programs”, *The Internet and Higher Education*, v. 6, n. 1, p. 1-16. [https://doi.org/10.1016/s1096-7516\(02\)00158-6](https://doi.org/10.1016/s1096-7516(02)00158-6).
- Siemens, G. and Baker, R. S. (2012) “Learning analytics and educational data mining”, In: *Proceedings of the 2nd LAK*, ACM, p. 252-254.
- Tinto, V. (1993) “Building community”, *Liberal Education*, v. 79, n. 4, p. 16-21.