

Learning Analytics para Moodle em uma arquitetura na nuvem: uma solução escalável para predição de risco acadêmico

Cristian Cechinel^{1,2,5}, Emanuel Marques Queiroga³, Tiago Thompsen Primo³,
Vinicius Faria Culmant Ramos¹, Roberto Muñoz⁴, Matheus Machado¹,
Mayara Stein¹, Juary Rocha², Thomas Michels Rodrigues⁶,
Henrique Lemos dos Santos², Giovani Maia Portelinha^{2,3}, Rafael Targino¹,
Valter Melgarejo¹, Juliano Almeida¹

¹Universidade Federal de Santa Catarina (UFSC)

²Universidade Federal de Pelotas (UFPel)

³Instituto Federal Sul-rio-grandense (IFSul)

⁴Universidad de Valparaíso - Chile (UV)

⁵CESAR School

⁶Universidade Regional de Blumenau (FURB)

emanuelqueiroga@gmail.com, tiago.primo@inf.ufpel.edu.br,

Abstract. *This paper delves into the development and enhancement of a Learning Analytics (LA) solution aimed at identifying patterns in student behavior within virtual learning environments (VLEs). This is achieved by generating predictive models that assess the risk of dropout and academic failures. Our proposed methodology incorporates a plugin within the institution's Moodle server. This plugin continually sends batches of data, capturing student interactions in the VLE, to a cloud-based platform. Subsequently, this data is processed and displayed on an interactive dashboard, providing predictions on students likely to face academic hurdles. The core aim of this solution is to present pivotal insights to educators and institutional leaders. This empowers them to detect early signs of potential academic issues and initiate timely interventions to improve educational outcomes.*

Resumo. *Este trabalho discute a implementação e aprimoramento de uma solução de Learning Analytics destinada à identificação de padrões de comportamento de estudantes em Ambientes Virtuais de Aprendizagem (AVA) e à criação de modelos preditivos para evasão e reprovação. A estratégia proposta engloba a instalação de um plugin no servidor Moodle da instituição, responsável pela coleta de informações relativas às interações dos estudantes no AVA. Estes dados são posteriormente processados em uma plataforma na nuvem e apresentados aos usuários através de um dashboard interativo. O principal objetivo desta solução é fornecer insights valiosos a educadores e administradores, possibilitando a detecção precoce de tendências de evasão ou reprovação e a tomada de medidas proativas para aprimorar os desempenhos acadêmicos.*

1. Introdução

Com a democratização do acesso a internet e aos dispositivos digitais, os processos de ensino e aprendizagem passam, gradativamente, por uma significativa transformação. Esse cenário é impulsionado pelas tecnologias da informação e comunicação, resultando na adoção crescente de Ambientes Virtuais de Aprendizagem (AVAs) [Siemens 2013, Siemens and Long 2011, Romero and Ventura 2020]. Esses ambientes oferecem flexibilidade e acessibilidade, permitindo que estudantes participem de cursos e atividades educacionais remotamente, rapidamente se tornando ferramentas de auxílio a aprendizagem nas diferentes modalidades de ensino contemporâneas.

Contudo, mesmo diante desse cenário, a evasão de estudantes emerge como um desafio considerável a ser superado, principalmente quando fazemos um recorte dos cursos na modalidade a distância (EAD). No Brasil, de acordo com o censo EAD 2020 [ABED 2022], as taxas de evasão nos cursos superiores a distância variam de 23% a 30%, enquanto que em outras modalidades é de aproximadamente 25%. Cabe ressaltar a importância da utilização dos ambientes virtuais de aprendizagem como ferramenta de apoio às modalidades de ensino, não sendo restrito ao EAD. Alguns estudos mostram como o comportamento dos estudantes no AVA pode revelar informações importantes sobre o risco de evasão e o desempenho dos estudantes em cursos presenciais [Queiroga et al. 2022, Baker and Inventado 2014].

Nesse sentido, é reconhecido o papel da identificação antecipada dos alunos em risco de evasão ou reprovação para que seja possível implementar medidas de intervenção adequadas para se tentar reverter esta situação [Manhães et al. 2011, Baker et al. 2011]. Portanto, a identificação precoce possibilita o aprimoramento dos processos de ensino, com a diminuição das taxas de evasão e retenção, e, por conseguinte, a promoção do êxito do processo educacional como um todo [Romero and Ventura 2020].

A partir dessa perspectiva, a *Learning Analytics* (LA) se apresenta como uma alternativa para geração de informação e conhecimento no processo educacional, a partir do tratamento dos dados gerados pelos AVAs [Baker and Inventado 2014]. No entanto, apesar do potencial da LA, ainda existe uma lacuna significativa na área de pesquisa no que diz respeito à implementação prática das teorias geradas, trazendo dificuldades para que estas teorias se convertam em práticas e atinjam o usuário final, seja ele o estudante, o professor ou a administração da organização [Cechinel et al. 2020].

Nesse contexto, este artigo explora a perspectiva de desenvolvimento e implementação de um serviço de LA para Moodle, que visa proporcionar a visualização para o professor dos resultados das previsões de risco dos estudantes, além de visualizações sobre as suas interações na disciplina. Esse *dashboard* tem o intuito de preencher a lacuna entre a teoria da LA e sua aplicação prática, oferecendo uma ferramenta eficaz para auxiliar os diversos atores envolvidos no processo educacional.

2. Revisão da literatura

A LA é uma área de pesquisa multidisciplinar, que se baseia em campos diversos, mas não restrito a eles, como Aprendizado de Máquina, Inteligência Artificial, Estatística e Visualização de Dados [Chatti et al. 2013, Baker and Inventado 2014, Clow 2012]. Essas disciplinas oferecem a base teórica e as ferramentas necessárias para a coleta, análise e

interpretação de dados educacionais, aprimorando a experiência de aprendizado. Além disso, a criação de modelos preditivos, através da identificação de padrões de comportamento dos alunos, busca personalizar o ensino e oferecer *insights* práticos aos educadores [Gómez-Pulido et al. 2020, Herodotou et al. 2019].

Pesquisadores têm apontado a contribuição central da LA na compreensão abrangente dos elementos formadores do aprendizado, analisando de maneira humanizada as variáveis que podem influenciar nas diferentes situações, como a conclusão de cursos ou o desempenho dos alunos em avaliações. Portanto, a LA dedica-se constantemente à identificação de padrões de comportamento das partes envolvidas no processo de aprendizagem, adotando uma abordagem holística [Ferguson 2012].

O ciclo de aprimoramento em LA é contínuo e não tem um ponto final predefinido. Dessa forma, a análise de aprendizagem está sempre em constante treinamento e reavaliação. Como resultado, a análise de aprendizagem está profundamente entrelaçada com áreas além da Mineração de Dados Educacionais (EDM), incluindo Inteligência de Negócios (BI) e Sistemas de Recomendação [Ferguson 2012]. Essa abordagem busca ampliar a compreensão do processo educacional e fortalecer a aplicação de estratégias eficazes para aprimorar a aprendizagem e alcançar resultados mais significativos.

A aplicação prática de Learning Analytics (LA) é promissora, mas enfrenta desafios significativos, tais como a complexidade dos dados educacionais, a carência de infraestrutura adequada e a resistência à inovação em instituições educacionais [Herodotou et al. 2017]. No entanto, diversos estudos evidenciam o crescente interesse de pesquisadores nesse campo. Estes estudos frequentemente se concentram na modelagem e extração de informações úteis para as instituições educacionais. Dentro deste âmbito, duas categorias de pesquisa se sobressaem: predição de desempenho e evasão, e visualização de dados. Ambas exploram dados oriundos de variadas fontes – desde sistemas acadêmicos até dados demográficos – visando compreender padrões de aprendizagem e identificar oportunidades de aprimoramento.

No campo da visualização de dados, as pesquisas concentram-se no desenvolvimento de técnicas específicas para painéis de visualização em LA, conhecidos como Learning Analytics Dashboards (LADs) [Romero and Ventura 2010, Saar et al. 2022]. Exemplos incluem abordagens para a visualização interativa de grandes conjuntos de dados educacionais, permitindo que os usuários explorem os padrões dos mesmos com diversos objetivos. Outra perspectiva é a visualização de dados multimodais, que integra diferentes fontes, como texto, imagens e vídeos, para oferecer uma representação mais rica dos dados educacionais [Sclater et al. 2016, Herodotou et al. 2019].

A próxima seção apresentará a arquitetura de uma solução prática de LA já implementada e em utilização, denominada de solução LANSE.

3. Arquitetura desenvolvida

A concepção da solução proposta baseou-se em um processo de design centrado no usuário, priorizando a usabilidade e a adaptabilidade. Para guiar o desenvolvimento da solução e validar suas características, foram conduzidas entrevistas com 31 *stakeholders* de janeiro a abril de 2022, abrangendo uma variedade de perfis, incluindo professores de educação superior e técnica, gestores universitários, especialistas em Educação

a Distância, profissionais da RNP e representantes do MEC. Essas entrevistas apontaram a necessidade da construção de uma solução com arquitetura em nuvem e uma característica importante: que requisitasse pouco trabalho das equipes técnicas das instituições clientes. Com o desenvolvimento de uma solução em nuvem, os clientes precisariam de um menor tempo dedicado a instalação e manutenção da solução, o que facilitaria a sua utilização e conseqüente escalabilidade.

A solução atual possui um plugin Moodle (bloco) do lado do cliente que é responsável por 1) recolher as interações que estão armazenadas nos logs do sistema e 2) enviar esses logs para serem pré-processados pelo serviço de LA. Para que os modelos de predição possam ser treinados e gerados para serem utilizados em um determinado curso, o *plugin* também deverá repassar informações relacionadas ao cronograma do mesmo (início e fim), além de permitir a seleção de cursos semelhantes que já tenham sido ministrados anteriormente e que sirvam como base para o treinamento e geração dos modelos de predição (essa última ainda em desenvolvimento).

Do ponto de vista puramente técnico, o *plugin* Moodle (bloco) possui um *pooling* que verifica novas informações de tempos em tempos e realiza a coleta de dados para processamento. Depois, este envia uma notificação via *push* para o serviço de LA na nuvem. O envio de *batches* é assíncrono, com envio diário e que são armazenados em uma fila. Do lado do serviço de LA na nuvem, são desenvolvidos os *scripts* para pré-processamento dos dados para o treinamento dos modelos de predição, para a execução desses modelos e a apresentação do *Dashboard* com as visualizações.

Na Figura 1, o módulo *enable* é responsável por controlar a lógica de habilitar e desabilitar o *dashboard* para uma disciplina específica. O módulo de segurança fica na frente da aplicação, sendo que todas as solicitações (*requests*) passam por ele para validar se os pedidos estão vindo de lugares autorizados. Para isso, inicialmente é utilizada a biblioteca de criptografia PyJWT¹. O módulo de processamento de logs é o responsável por fazer as lógicas de processamento para os modelos de aprendizado de máquina e também para os gráficos. Nesse módulo, estão sendo utilizadas principalmente as bibliotecas *pandas*² e *scikit-learn*³ do Python. O módulo de *machine learning* é responsável por treinar os modelos de predição e salvar os resultados no banco de dados, para posteriormente serem apresentados no *frontend*. Ainda, o módulo de visualizações entrega os dados de visualização no formato correto para o *frontend* e o módulo de logs é responsável por salvar os dados recebidos do *plugin* Moodle.

O processo inicia quando os usuários professores habilitam o *plugin* dentro de seus cursos no Moodle (Figura 2). Uma vez ativado, o *plugin* começa a coleta sistemática dos dados, capturando informações detalhadas sobre os acessos dos estudantes ao ambiente virtual de aprendizagem. O professor pode acessar a solução na nuvem clicando em *Access Dashboard*.

Os dados coletados são enviados de forma segura para um servidor na nuvem, onde ocorre o processamento centralizado. Nesse servidor, técnicas de processamento de dados em larga escala são empregadas para transformar os dados brutos em informações

¹<https://pyjwt.readthedocs.io/en/stable/>

²<https://pandas.pydata.org/>

³<https://scikit-learn.org/stable/>

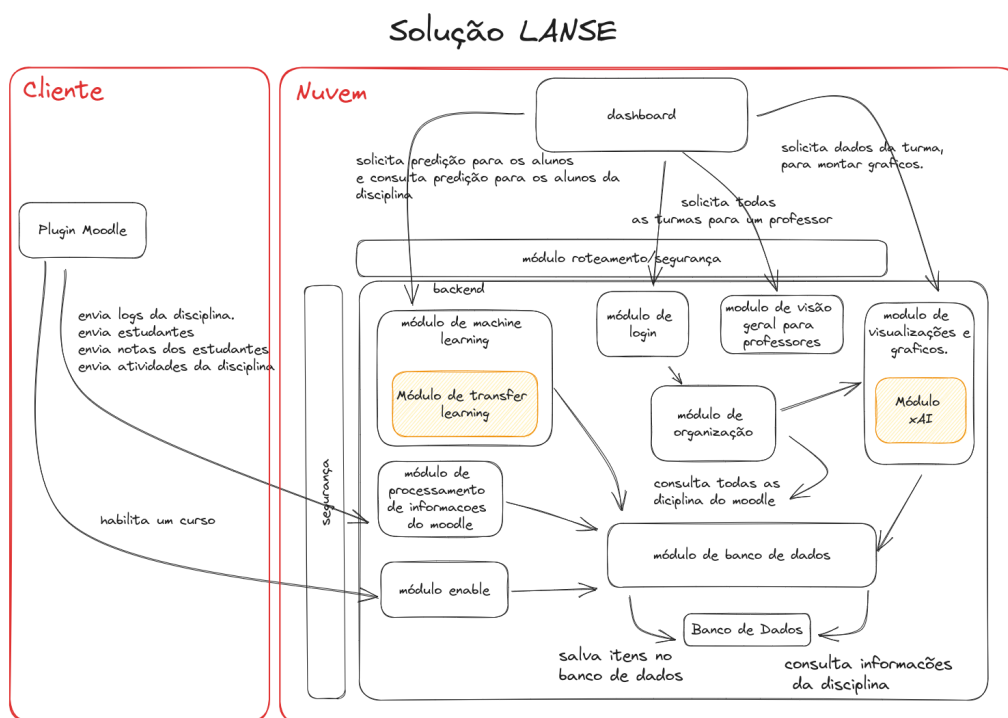


Figure 1. Arquitetura da solução LANSE

Demo (20231)

Painel / Meus cursos / Demo (20231)

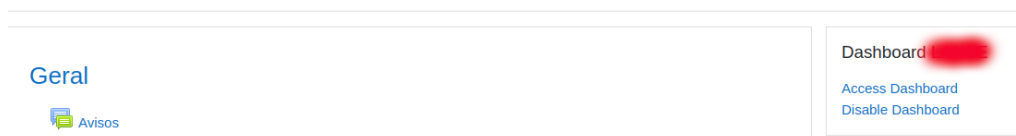


Figure 2. Acesso ao dashboard na nuvem

úteis e mensuráveis. Os dados são organizados em categorias relevantes, como tipos de atividades realizadas e interações específicas, bem como em uma série de dados derivados, permitindo uma análise mais precisa e detalhada. A arquitetura oferece flexibilidade para analisar os dados em diferentes níveis de granularidade, incluindo acessos diários, semanais e mensais.

Após o envio inicial, os dados ficam disponíveis em um *dashboard* de visualização de dados, que exibe de maneira interativa as informações processadas. Os usuários têm acesso a métricas detalhadas sobre os padrões de acesso dos estudantes, bem como seus níveis de engajamento em diferentes momentos, atividades e recursos educacionais. Ainda, o *dashboard* sinaliza possíveis riscos de evasão, permitindo que os educadores possam intervir para oferecer suporte adequado, e, por exemplo, possam enviar mensagens aos estudantes. A Figura 3 apresenta a tela inicial em que o professor da disciplina no Moodle pode ter uma visualização geral dos percentuais de acesso e risco dos estudantes, assim como também a listagem de seus estudantes, sua situação de risco e a confiança do modelo de predição com relação ao risco calculado.



Figure 3. Visão geral

A predição é realizada de maneira assíncrona com uma arquitetura contendo sistema de eventos de mensageria garantindo a independência do sistema e não precisando recalculando solicitações já realizadas. O sistema é capaz de rastrear as mensagens já enviadas (e também as que falharam), possibilitando detectar *bugs* mais facilmente uma vez que é possível replicar as mensagens que entraram no sistema. As predições são calculadas e atualizadas semanalmente (ver Figura 4, permitindo ao professor acompanhar mudanças com relação ao risco de seus estudantes.



Figure 4. Risco ao longo das semanas

Além das informações iniciais, também são oferecidas outras visualizações relacionadas às interações com a disciplina, tais como as quantidades de interações totais de cada estudante individualmente e acumuladas ao longo das semanas (ver Figuras 5 e 6).

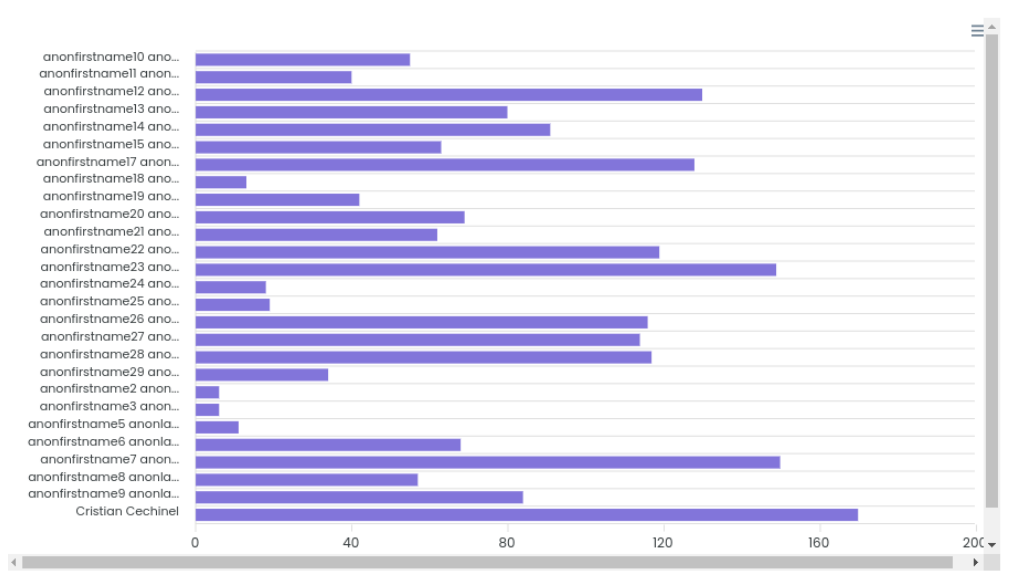


Figure 5. Total de acessos dos estudantes

3.1. Sobre os modelos de predição

Com o objetivo de contornar o problema de transferência e escalabilidade de uma solução de predição de risco acadêmico, nos últimos anos, o presente grupo de pesquisa tem desenvolvido modelos de predição baseados na contagem simples das interações dos acadêmicos para cursos disponibilizados em ambientes virtuais de aprendizagem (Moodle). A viabilidade dessa estratégia de utilizar apenas a quantidade de interações dos acadêmicos com os cursos para gerar predições com desempenho razoáveis já foi confirmada em trabalhos prévios dos proponentes [Macarini et al. 2019, Machado et al. 2018, Detoni et al. 2016]. Nesses trabalhos, também foram testadas a derivação de atributos que levam em conta a quantidade de interações de estudantes, bem como outros atributos derivados. A utilização destes atributos nos modelos permitiram que os mesmos alcançassem desempenhos altos, apresentando taxas, nas primeiras semanas de curso (até a 4a semana de aula), com AUCROC de, aproximadamente, 70% a 80%. Outros testes, com modelos baseados em contagem de interações também confirmam a possibilidade de sua utilização para a predição da evasão [Queiroga et al. 2020]. Esses modelos oferecem a facilidade de serem transferidos para um produto que possa ser utilizado por diferentes instituições e cenários, permitindo a escalabilidade dos modelos e funcionando independentemente do desenho instrucional de cada curso, uma vez que os parâmetros de entrada dos modelos são sempre os mesmos. Essa abordagem é a que está atualmente implementada na solução.

4. Considerações finais

O estudo em questão investigou a criação e aperfeiçoamento de uma ferramenta de LA, focada na identificação de comportamentos de estudantes em Ambientes Virtuais de Aprendizagem (AVAs) e na formulação de modelos que predizem a evasão. Sua arquitetura possibilitou a coleta, processamento e exibição de dados educacionais, oferecendo *insights* preciosos para educadores e gestores, melhorando a tomada de decisões e impulsionando resultados acadêmicos.

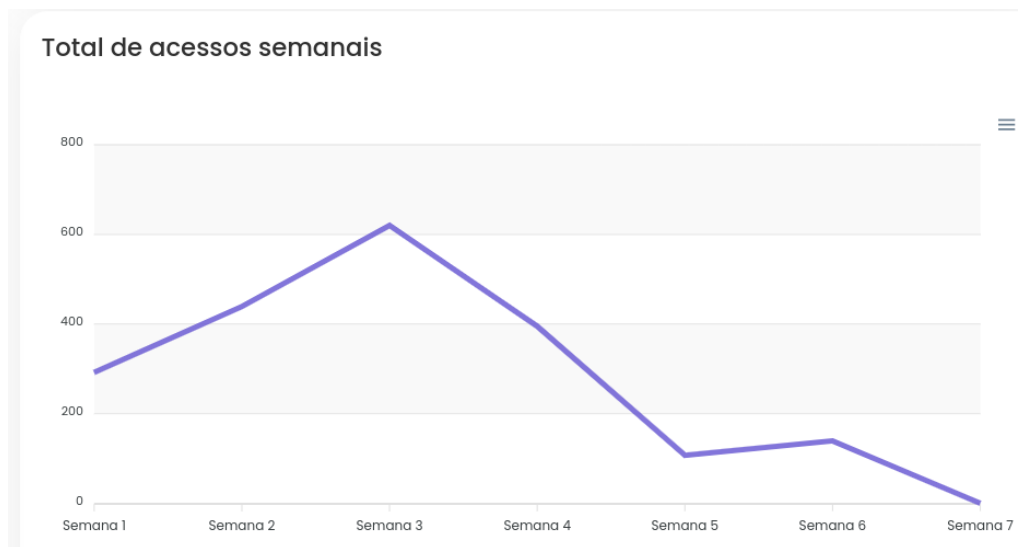


Figure 6. Totais de acessos semanais

Atualmente, esta ferramenta é usada em algumas disciplinas de cursos superiores à distância em duas instituições. O *feedback* e os efeitos desse uso serão abordados em trabalhos futuros.

Ao implementar a ferramenta, deparamo-nos com desafios, especialmente em relação à privacidade e conformidade com a Lei Geral de Proteção de Dados (LGPD). A preocupação com a proteção de dados sensíveis exigiu a aderência rigorosa a regulamentações e uma comunicação clara sobre as práticas de privacidade.

Outra barreira foi o tratamento e armazenamento de grandes volumes de dados. Era crucial gerir os recursos computacionais, assegurar escalabilidade e manter a integridade dos dados para a robustez da ferramenta. A necessidade de processamento em tempo real e de garantir segurança das informações nos levou a adotar tecnologias de ponta e práticas de engenharia sólidas.

Fomos ainda desafiados pela falta de dados históricos em algumas instituições, essenciais para criar modelos preditivos. Dada a importância dos dados históricos para treinar algoritmos, essa ausência afetou a precisão dos modelos, levando-nos a considerar métodos alternativos.

Como próximo passo, estamos considerando o uso de técnicas de *Transfer Learning* para superar a falta de dados históricos. Essa técnica permitiria aproveitar conhecimentos de modelos de outras instituições, ajustando-os a contextos específicos. A ideia é viabilizar modelos preditivos mesmo com escassez de dados.

Adicionalmente, vislumbramos a expansão das funcionalidades do *dashboard*, incorporando análises mais detalhadas e representações visuais avançadas. Isso proporcionaria uma visão mais completa dos padrões de comportamento dos estudantes, auxiliando na identificação de áreas para intervenção e melhoria.

Agradecimentos

Este trabalho foi financiado pela Rede Nacional de Ensino e Pesquisa (RNP) por meio de seu Programa de PD&I e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) por meio do Projeto IA2 – Empreendedorismo de Base Tecnológica em Inteligência Artificial (proc. 409633/2022-4). Cristian Cechinel foi parcialmente financiado pelo CNPq [DT-2 - Bolsa de Produtividade em Desenvolvimento Tecnológico e Extensão Inovadora, proc.305731/2021-1].

References

- ABED, A. B. D. E. A. D. (2022). *Censo EAD.BR: Analytic Report of Distance Learning in Brazil 2020. InterSaberes. (Obra original publicada em 2022).*
- Baker, R., Isotani, S., and Carvalho, A. (2011). Mineração de Dados Educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, 19(02):03.
- Baker, R. S. and Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics*, pages 61–75. Springer.
- Cechinel, C., Ochoa, X., Lemos dos Santos, H., Carvalho Nunes, J. B., Rodés, V., and Marques Queiroga, E. (2020). Mapping learning analytics initiatives in latin america. *British Journal of Educational Technology*, 51(4):892–914.
- Chatti, M. A., Dyckhoff, A. L., Schroeder, U., and Thüs, H. (2013). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6):318–331.
- Clow, D. (2012). The learning analytics cycle: closing the loop effectively.
- Detoni, D., Cechinel, C., Matsumura, R. A., and Brauner, D. F. (2016). Learning to identify at-risk students in distance education using interaction counts. *Revista de Informática Teórica e Aplicada*, 23(2):124–140.
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6):304–317.
- Gómez-Pulido, J. A., Park, Y., and Soto, R. (2020). Advanced techniques in the analysis and prediction of students’ behaviour in technology-enhanced learning contexts.
- Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., Hlosta, M., and Naydenova, G. (2017). Implementing predictive learning analytics on a large scale: the teacher’s perspective. In *Proceedings of the seventh international learning analytics & knowledge conference*, pages 267–271.
- Herodotou, C., Rienties, B., Verdin, B., and Boroowa, A. (2019). Predictive learning analytics ‘at scale’: Towards guidelines to successful implementation in higher education based on the case of the open university uk. *Journal of Learning Analytics*, pages In-Press.
- Macarini, B., Antonio, L., Cechinel, C., Batista Machado, M. F., Faria Culmant Ramos, V., and Munoz, R. (2019). Predicting students success in blended learning—evaluating different interactions inside learning management systems. *Applied Sciences*, 9(24):5523.

- Machado, M., Cechinel, C. C., and Ramos, V. (2018). Comparação de diferentes configurações de bases de dados para a identificação precoce do risco de reprovação: o caso de uma disciplina semipresencial de algoritmos e programação. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, 29(1):1503.
- Manhães, L. M. B., Cruz, S. d., Costa, R. J. M., Zavaleta, J., and Zimbrão, G. (2011). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. *Anais do XXII SBIE-XVII WIE, Aracaju*.
- Queiroga, E. M., Lopes, J. L., Kappel, K., Aguiar, M., Araújo, R. M., Munoz, R., Villarroel, R., and Cechinel, C. (2020). A learning analytics approach to identify students at risk of dropout: A case study with a technical distance education course. *Applied Sciences*, 10(11).
- Queiroga, E. M., Paragarino, V. R., Casas, A. P., Primo, T. T., Munoz, R., Ramos, V. C., and Cechinel, C. (2022). Experimenting learning analytics and educational data mining in different educational contexts and levels. In *2022 XVII Latin American Conference on Learning Technologies (LACLO)*, pages 1–9. IEEE.
- Romero, C. and Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6):601–618.
- Romero, C. and Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3):e1355.
- Saar, M., Rodríguez-Triana, M. J., and Santos, L. P. P. (2022). Towards data-informed teaching practice:: A model for integrating analytics with teacher inquiry. *Journal of Learning Analytics*, 9(3):88–103.
- Sclater, N., Peasgood, A., and Mullan, J. (2016). Learning analytics in higher education. *London: Jisc. Accessed February*, 8(2017):176.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10):1380–1400.
- Siemens, G. and Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5):30.