

SABIA: Uma Plataforma para Auxiliar a Gestão Baseada em Evidências nas Instituições de Ensino Superior

Ebony Marques¹, Lhaislla Cavalcanti¹, Tarcísio Barbosa¹ e Gabriel Alves¹

¹Universidade Federal Rural de Pernambuco (UFRPE)
Recife — PE — Brasil

{ebony.marquesr, lhaislla.cavalcanti,
tarcisio.bcosta, gabriel.alves}@ufrpe.br

Abstract. *Evidence-based management gains prominence in organizational decision-making, promoting a culture that values decision-making based on data, best practices, and research. Business Intelligence (BI) and Analytics are disciplines that operate from data collection to analysis and prediction to support this type of management. The System of Academic Business Intelligence and Analytics (SABIA) is a platform that combines these disciplines to assist evidence-based management in Higher Education Institutions. Covering various academic dimensions and key performance indicators, SABIA aids in understanding the current situation and making predictions related to students and courses in Higher Education Institutions.*

Resumo. *A gestão baseada em evidências ganha destaque na tomada de decisões organizacionais, promovendo uma cultura que valoriza a fundamentação de escolhas por meio de dados, boas práticas e pesquisas. O Business Intelligence (BI) e o Analytics são disciplinas que atuam desde a coleta até a análise de dados e realização de previsões visando auxiliar a gerência. O System of Academic Business Intelligence and Analytics (SABIA) é uma plataforma que combina essas disciplinas a fim de auxiliar a gestão baseada em evidências em Instituições de Ensino Superior (IES). Cobrindo várias dimensões acadêmicas e indicadores-chave de desempenho, o SABIA auxilia a compreensão atual e a realização de previsões relacionadas aos alunos e cursos nas IES.*

1. Introdução

A gestão baseada em evidências procura estabelecer uma cultura organizacional em que as decisões são tomadas com base em dados, boas práticas e pesquisas. Esse modelo de gestão se contrapõe a modelos baseados em empirismos, valendo-se de sistemas que permitam a obtenção de informações com base nos dados que viabilizam a extração de *insights* fundamentais no âmbito das decisões organizacionais. No contexto educacional, as abordagens de *Learning Analytics* (LA) e *Academic Analytics* (AA) desempenham papel central no apoio à decisão [Ferguson 2012, Siemens 2013, Romero and Ventura 2020].

Se por um lado ambas abordagens tratam da coleta, análise e predição com base em dados educacionais, por outro lado elas divergem um pouco quanto ao foco. Enquanto o *Learning Analytics* foca especialmente o desempenho dos estudantes e as interações dentro da sala de aula, o *Academic Analytics* expande o foco para a esfera institucional. Essa mudança de foco impacta desde os sistemas utilizados para a coleta dos dados até os

modelos e métricas empregados. Contudo, há de se salientar que ambas também possuem muitos elementos e objetos de análise e predição em comum, como a predição da evasão estudantil [Krumm et al. 2018, Romero and Ventura 2020]. Essas perspectivas de análise tornam-se cruciais e complementares ao considerar, por exemplo, o ingresso de estudantes no ensino superior e o seu percurso na vida acadêmica [Viana et al. 2022].

O *Business Intelligence* (BI) engloba ferramentas e processos que atuam desde a coleta até a análise dos dados, geralmente estruturados, com o objetivo de fornecer insumos para a gestão baseada em dados. Enquanto o BI procura entender a situação anterior e atual da empresa, o *Analytics* procura realizar predições com base nesses dados, que podem ou não ser estruturados, também com o objetivo de contribuir para uma gestão baseada em dados [Chen et al. 2012].

Este trabalho apresenta o *System of Academic Business Intelligence and Analytics* (SABIA), que agrega conceitos de *Learning Analytics*, *Academic Analytics* e *Business Intelligence* com o objetivo de dar suporte à gestão baseada em evidências em Instituições de Ensino Superior (IES). Ele fornece painéis que focam diferentes dimensões acadêmicas, que vão desde a visão geral da instituição até o detalhamento das disciplinas dos cursos. Além dos indicadores-chave de desempenho tradicionais da área de educação, como o Índice Geral de Cursos (IGC) e o Conceito Preliminar de Curso (CPC), outros indicadores, como a Taxa de Sucesso de Disciplina (TSD), também foram elaborados para auxiliar a análise de problemas relacionados à retenção.

2. Trabalhos Relacionados

Esta seção apresenta alguns trabalhos relacionados nas áreas de *Learning Analytics* (LA), *Academic Analytics* (AA) e *Business Intelligence* (BI) no âmbito educacional. O trabalho de [Palomino et al. 2022] trata da importância das plataformas de análise de dados educacionais para os gestores. No estudo em questão, enfatiza-se a relevância da visualização de dados no suporte à tomada de decisão. Apesar de ser aplicado para supervisão das metas do Plano Nacional de Educação (PNE) no Brasil, o trabalho mostra que plataformas de análise com boas visualizações de dados também podem ser empregadas no contexto da gestão de instituições de ensino superior.

A utilização de *Learning Analytics* (LA) é bastante comum em Ambientes Virtuais de Aprendizagem (AVA) [Portela et al. 2020, Padilha and Souza 2017] e especialmente útil em disciplinas que aplicam *Problem Based Learning* (PBL) [Aguiar et al. 2021]. O LA busca aferir métricas que vão desde o acesso a objetos de aprendizagem até a hora de acesso ao sistema, a fim de realizar predições sobre o desempenho dos estudantes. Assim, o LA contribui para o aprimoramento do processo de ensino-aprendizagem dos estudantes, melhorando ainda o relacionamento entre professores e alunos.

Os estudos tratados em [Mesquita et al. 2021] e [Ferreira and Andrade 2013] ressaltam a importância do *Academic Analytics* (AA) na gestão eficiente nas instituições de ensino superior, especialmente em face do aumento do volume de dados relacionados à educação. Em especial, o trabalho de [Mesquita et al. 2021] trata de uma revisão sistemática da literatura na área de AA, visando identificar os dados acadêmicos comumente empregados para diagnosticar os índices de sucesso na graduação e as abordagens utilizadas para aprimorar o sucesso e reduzir o insucesso estudantil.

Trabalhos como os apresentados em [Júnior et al. 2022] e [Melo and Sousa 2022]

tratam do desenvolvimento e do emprego de soluções de BI como recursos para fundamentar a tomada de decisão no contexto acadêmico. O trabalho de [Júnior et al. 2022] compartilha as experiências obtidas durante o desenvolvimento de uma plataforma que visa reduzir a evasão estudantil em uma instituição de ensino superior, enquanto a pesquisa de [Melo and Sousa 2022] propõe um modelo para implementação de uma solução genérica para o auxílio à gestão acadêmica. Os estudos usaram técnicas e ferramentas de BI para fornecer suporte às decisões estratégicas de instituições de ensino. A partir dos resultados percebidos, os trabalhos defendem que o BI é um importante instrumento para o auxílio à gestão acadêmica.

Ao integrar as abordagens de LA, AA e BI, a plataforma apresentada neste trabalho vai ao encontro do que tem sido buscado em trabalhos relacionados executados em outras instituições de ensino superior brasileiras, contribuindo diretamente para a gestão de IES. Além de identificar situações que geram maior risco de evasão, os resultados obtidos neste trabalho podem contribuir diretamente para uma cultura de gerenciamento baseado em evidências. A plataforma busca apoiar coordenadores de cursos e gestores acadêmicos no monitoramento de indicadores-chave de desempenho, auxiliando a elaboração de políticas institucionais que, em última análise, contribuam para a redução da evasão estudantil.

3. Método

O SABIA foi desenvolvido seguindo os processos utilizados na construção de sistemas de BI, por meio da linguagem de programação Python¹ e de bibliotecas como Streamlit² e Scikit-Learn³. Cada painel da plataforma utiliza os dados de um *data mart* único, tornando seu processo de carregamento mais leve e simplificando o próprio desenvolvimento do painel. Esses *data marts* foram elaborados com a observação dos princípios de *pipelines de dados ETL* [Raj et al. 2020]. Uma vez que a menor granularidade se refere ao desempenho dos estudantes em disciplinas, não há necessidade do uso de um *pipeline* de dados *real-time*, sendo o processo de ETL executado em lote para carregar os *data marts* pelo menos uma vez por semestre. As principais etapas da elaboração da plataforma de BI são:

- **Ingestão dos Dados:** São identificadas as diferentes fontes e coletados os dados da instituição de ensino superior. Dentre as fontes, citam-se o Censo da Educação Superior [INEP 2023], os resultados do ENADE e os seus sistemas de gestão, além de fontes internas da instituição, como planilhas.
- **Limpeza dos Dados:** Os dados são processados a fim de tratar inconsistências. Dados duplicados, redundâncias e erros são tratados, garantindo a qualidade e consistência das informações.
- **Transformação dos Dados:** Os dados são preparados para o carregamento nos *data marts*. Os dados de uma mesma fonte podem ser incorporados em diferentes *data marts*. É nesta etapa que são realizadas transformações nos dados a fim de compatibilizar diferentes fontes ou versões de uma mesma fonte. Por exemplo, caso ocorra uma mudança na forma de disponibilizar os dados do Censo da

¹<https://www.python.org/>

²<https://streamlit.io/>

³<https://scikit-learn.org/>

Educação, é nesta etapa que as adaptações serão carregadas. Agregações, filtrações, renomeação de campos, codificação de valores ou criação de campos calculados podem ser realizadas antes do carregamento dos dados nos *data mart*.

- **Carregamento dos Dados:** Os dados já preparados são carregados em cada *data mart* específico.
- **Compartilhamento de Resultados:** Os resultados são compartilhados em painéis contendo opções de filtragem, tabelas, gráficos e mapas de calor, a fim de promover melhoria contínua dos cursos e da instituição.

A plataforma foi estruturada de forma a dar suporte ao processo de adoção de AA e LA [Schneider et al. 2022]. Portanto, ela foi desenvolvida em alinhamento com demandas de diferentes setores da IES interessados na gestão acadêmica. Também há textos explicativos sobre o uso na própria plataforma. Assim, as informações fornecidas pelo sistema podem ser utilizadas como insumo para a elaboração de ações institucionais, como no caso do cálculo da quantidade de vagas em processos seletivos extras.

O acesso público do SABIA⁴ apresenta os dados gerais da instituição, enquanto o acesso privado por meio de *login* permite a visualização de informações mais específicas do curso ou mesmo do fluxo dos estudantes ao longo dos períodos. Em ambas formas de acesso, diferentes opções de filtragem são disponibilizadas para o usuário, permitindo-lhe focar dimensões específicas de seu interesse. Algumas das funcionalidades da plataforma são apresentadas na Seção 4.

4. Resultados e Discussões

Esta seção apresenta algumas das funcionalidades e possibilidades de análises realizadas a partir da plataforma proposta. O SABIA encontra-se em pleno funcionamento em uma IES pública federal, sendo utilizado por gestores tanto para o acompanhamento dos cursos quanto para determinadas ações, como o cálculo de vagas a serem disponibilizadas em processos de seleção específicos. A fim de facilitar o entendimento de algumas análises que podem ser realizadas, os dados apresentados nesta seção focam um curso da área de computação da instituição.

4.1. Informações Institucionais

O primeiro painel da plataforma é disponibilizado publicamente e fornece informações relacionadas aos indicadores institucionais. Alguns desses indicadores são os indicadores de qualidade do INEP (ENADE, CPC, IDD e IGC) e a Taxa de Sucesso da Graduação (TSG). Também são apresentadas informações como quantidades de campus e de estudantes, entre outras. A Figura 1 apresenta alguns trechos do painel em questão, com as opções de filtragem (à esquerda) e os principais indicadores (à direita). Os controles referentes ao *Período Letivo* e o *Período Letivo de Referência*, à esquerda da imagem que exibe os indicadores, são utilizados para alterar os valores apresentados no centro e no canto inferior direito dos *cards*, respectivamente. A variação entre esses valores também é apresentada no centro do *card*, assumindo a cor verde, caso seja positiva, ou vermelha, caso seja negativa.

Essa página também exibe gráficos com o histórico dos indicadores e com os totais de estudantes que se formaram e se evadiram, trancamentos de semestres e total

⁴<https://ufrpe-odg.github.io/site/sabia.html>



Figura 1. Filtros e indicadores-chave de desempenho institucionais do painel principal. As opções de filtragem alteram automaticamente os indicadores e gráficos desse painel.

de matrículas. Os mapas de calor apresentados na Figura 2 mostram o detalhamento por ano e por curso dos totais de formados e evadidos. Os dados foram obtidos com o ano letivo de 2022 em andamento, portanto é possível que existam bem menos formados nesse ano. Já o mapa de evadidos apresenta um dado que requer atenção e que reflete um comportamento geral nos cursos de graduação após a pandemia de COVID-19. O tom mais escuro de vermelho no ano de 2021 em praticamente todos os cursos aponta para uma maior evasão nesse ano. Essa situação precisa ser monitorada e, caso se mostre como uma tendência nos anos seguintes, políticas de acesso e permanência precisam ser elaboradas para reverter a evasão.

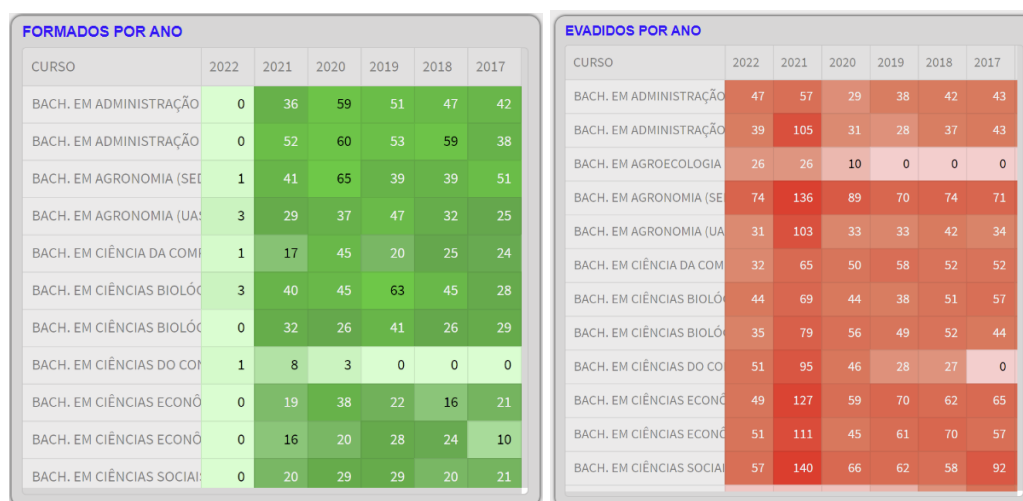


Figura 2. Mapas de calor relativos à quantidade de formados e evadidos, detalhados por curso e por ano.

4.2. Informações do Curso

Apesar de ser possível filtrar as informações de quaisquer painéis para um curso específico, alguns painéis agregam informações que são especialmente úteis para os gestores

dos cursos. Esses painéis apresentam *cards* e gráficos com informações similares àquelas apresentadas no painel principal, contudo pré-filtradas apenas para o curso selecionado. Além disso, outras informações que focam as disciplinas do curso são apresentadas.

O painel do curso permite a seleção de um *Período Letivo* e um *Período Letivo de Referência* a fim de comparar a variação de diferentes indicadores, como apresentado na Figura 3. É possível observar que a disciplina que mais aprovou no período 2019.1 foi a de Introdução a Programação, uma vez que os dados estão ordenados do maior para o menor para esse período. Apesar disso, essa também é a segunda disciplina que mais reprovou, como visto na imagem ao centro, atrás apenas da disciplina de Matemática Discreta. Como são disciplinas dos períodos iniciais do curso, os altos valores são esperados.

O indicador de taxa de sucesso calcula a razão entre os estudantes aprovados e matriculados em uma disciplina e foi elaborado a fim de identificar disciplinas com uma alta taxa de reprovação, independentemente do período em que são lecionadas. Na imagem à direita, a coluna *DIFF*(%) foi utilizada para a ordenação do menor para o maior. Assim, é possível observar as disciplinas que diminuíram a sua taxa de sucesso em 2019 com relação ao mesmo período de 2018. Com isso, o gestor pode observar e acompanhar essas variações, que podem ser causadas por diferentes fatores, como mudanças na metodologia de ensino.

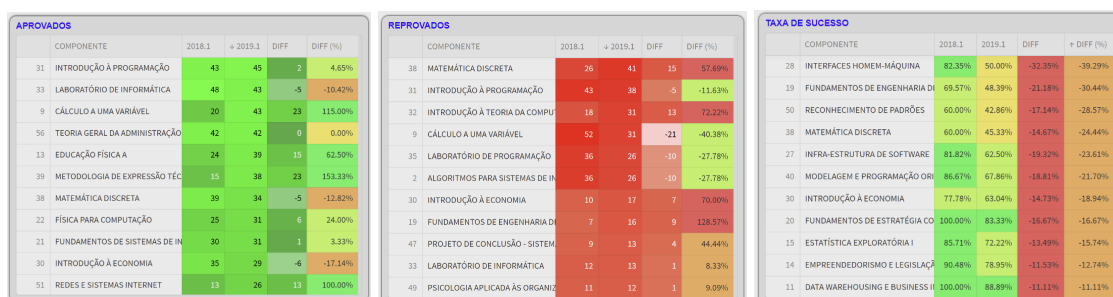


Figura 3. Tabelas comparativas com o total de aprovados, total de reprovados e a taxa de sucesso nas disciplinas de um curso da área de computação.

4.3. Informações das Disciplinas

A Figura 4 apresenta os gráficos de caixa e de bolha para as disciplinas de um curso da área de computação, para o período de 2010 a 2022. O diagrama de caixas exibe, para cada período, os valores relacionados à taxa de sucesso das disciplinas do curso. Além da taxa de sucesso, é possível selecionar outras métricas, como a média final dos aprovados e a quantidade de reprovações. Esse painel ainda possibilita a seleção de apenas um subconjunto de disciplinas, permitindo que sejam avaliadas, por exemplo, somente disciplinas obrigatórias ou de uma área de conhecimento específica.

Considerando o gráfico da Figura 4, observa-se que houve uma queda considerável na mediana da taxa de sucesso a partir do período 2020.1, que foi realizado após a pandemia de COVID-19. Observando o histórico da mediana, é possível inferir que, enquanto metade das disciplinas aprovavam mais de 75% dos estudantes, após a pandemia, essa mesma quantidade de disciplinas passou a aprovar apenas 60% a 65% dos estudantes. Ainda é possível notar que, a partir de 2020.1, o terceiro quartil possui aproximadamente o mesmo valor da mediana dos períodos anteriores. Ou seja, enquanto apenas 25% das

disciplinas aprovam mais de 75% dos estudantes após a pandemia de COVID-19, 50% das disciplinas aprovavam a mesma quantidade de estudantes antes desse período. Esse gráfico ainda permite observar os *outliers*, que indicam disciplinas específicas com valores discrepantes com relação às demais disciplinas do período.



Figura 4. Gráficos de caixa e de bolha para as disciplinas, filtrando apenas as disciplinas de um curso da área de computação entre 2010 e 2022.

O gráfico de bolhas exibido na Figura 4 permite que o usuário selecione as variáveis que serão apresentadas no eixo X, no eixo Y, no tamanho da bolha e na cor da bolha. Na imagem, o eixo X apresenta a média final, que substitui por 0 (zero) a nota dos estudantes reprovados por falta. A média final dos aprovados, que considera apenas os estudantes aprovados no cálculo da média, é exibida no eixo Y. O tamanho da bolha representa a quantidade de estudantes matriculados na disciplina, portanto, quanto maior a bolha, maior a quantidade de estudantes que cursaram a disciplina. Por fim, a escala de cor varia do vermelho ao verde e está associada à taxa de sucesso da disciplina. Essa imagem mostra os resultados considerando todas as turmas e estudantes que cursaram a disciplinas desde 2010. Na imagem, é possível observar que, ao passar o ponteiro do *mouse* sobre a bolha, são exibidas informações sobre a disciplina. Esse gráfico mostra que a disciplina Cálculo NI, que substituiu a disciplina de Cálculo a uma Variável, possui uma Taxa de

Sucesso de apenas 21,67% e teve 67 matrículas. Apesar disso, essa disciplina possui uma média relativamente alta (8,6), quando considerados apenas os estudantes aprovados.

4.4. Progresso Estudantil

O acompanhamento do progresso dos estudantes ao longo de sua formação acadêmica, representada por meio da relação entre o estudante e a instituição de ensino, é essencial para compreender a situação estudantil nos cursos. O gráfico Sankey (Figura 5) apresenta o progresso dos estudantes considerando a sua duração de vínculo com o curso e sua situação final e atual (matriculado, evadido, ou formado). Além desse gráfico, o SABIA também possui gráficos específicos para a análise de sobrevivência [dos Santos et al. 2022] dos estudantes, permitindo a comparação de atributos como curso, forma de ingresso e período de ingresso.

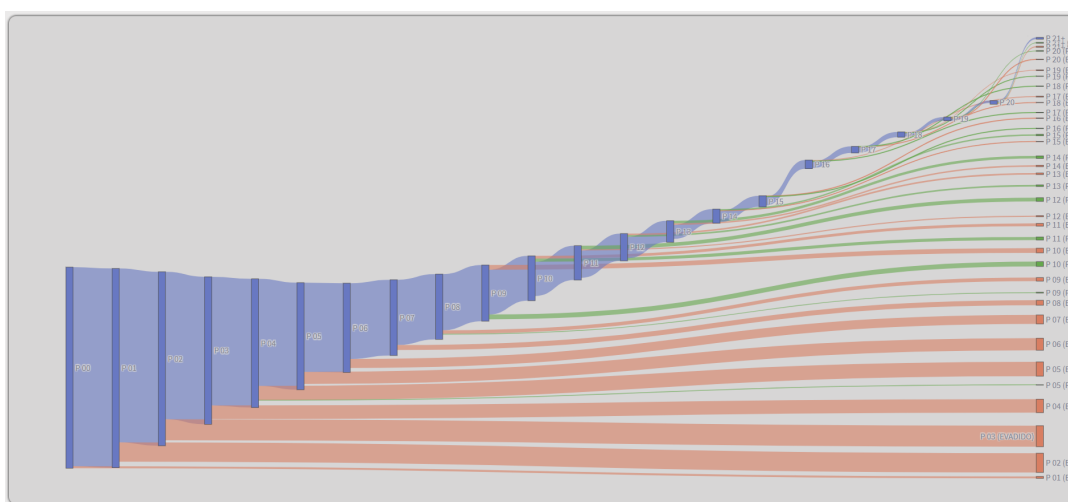


Figura 5. Progresso dos estudantes em um curso da área de computação.

Nesse gráfico, a transição entre um período de vínculo e outro é representada por uma cor que indica se o estudante permaneceu vinculado (azul), se ele se formou (verde) ou se ele se evadiu (vermelho). A espessura da faixa e da linha vertical é proporcional à quantidade de estudantes que estavam naquela situação. Assim, é possível notar que, entre as faixas vermelhas, as que estão marcadas com P01 e P02 são as mais grossas, indicando que a maior parte da evasão ocorre quando os estudantes completam 2 e 3 períodos de vínculo. Além disso, essa evasão parece permanecer alta do terceiro até o sétimo período, quando começa a diminuir.

5. Considerações Finais

Este trabalho trata de uma plataforma que exibe indicadores de desempenho de uma instituição, cursos, disciplinas e estudantes disponibilizados em painéis dinâmicos, aplicando conceitos de *Business Intelligence*, *Academic Analytics* e *Learning Analytics*. Esses painéis utilizam *cards*, gráficos, tabelas e mapas de calor a fim de converter os dados em informações para o gestores. O SABIA já se encontra em uso por uma instituição de ensino superior pública federal, contribuindo para a gestão de seus cursos. Outras funcionalidades, como a análise de risco de evasão dos estudantes e a sugestão de disciplinas a serem cursadas com base no perfil do estudante, serão desenvolvidas em trabalhos futuros.

Referências

- Aguiar, F., Mello, R. F., Furtado, A. P. C., Nascimento, A., and Miranda, P. (2021). Learning analytics e problem-based learning: Mapeamento sistemático da literatura. In *Anais do VI Congresso sobre Tecnologias na Educação*, pages 110–119. SBC.
- Chen, H., Chiang, R. H., and Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, pages 1165–1188.
- dos Santos, J. F., de Sousa, J. D. A., Mello, R. F., Cristino, C. T., and Alves, G. (2022). Análise da evasão e do impacto da retenção estudantil no ensino superior com cadeias de markov absorventes. In *Anais do XXI Workshop em Desempenho de Sistemas Computacionais e de Comunicação*, pages 49–59. SBC.
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5-6):304–317.
- Ferreira, S. A. and Andrade, A. M. V. d. (2013). Academic analytics na prática: agregar e analisar múltiplas fontes de dados. *Revista Portuguesa de Investigação Educacional*, 13:235–260.
- INEP (2023). Instituto nacional de estudos e pesquisas educacionais anísio teixeira. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior>. Acesso em: 09 junho 2023.
- Júnior, O. d. G. F., Carvalho, V. D. H. d., Barros, P. A. M., and Braga, M. d. M. (2022). Uma experiência com business intelligence para apoiar a gestão acadêmica em uma universidade federal brasileira. *Revista Ibérica de Sistemas e Tecnologias de Informação*, 46:5–20.
- Krumm, A., Means, B., and Bienkowski, M. (2018). *Learning analytics goes to school: A collaborative approach to improving education*. Routledge.
- Melo, A. I. R. d. and Sousa, R. R. d. (2022). Gestão de instituições de ensino: uma proposta simples de business intelligence para melhoria do processo de ensino-aprendizagem management of educational institutions: a simple business intelligence proposal to improve the teaching-learning process. *Brazilian Journal of Development*, 8(6):47048–47061.
- Mesquita, J. L. d., Sousa, R. R. d., Nascimento, S. M., and Souza, T. F. d. (2021). Academic analytics como apoio ao sucesso na graduação: uma revisão sistemática da literatura academic analytics to support undergraduate success: a systematic review of the literature. *Brazilian Journal of Development*, 7(10):99882–99897.
- Padilha, T. and Souza, I. (2017). Uma experiência do uso de ferramentas de learning analytics para análise de interações. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*.
- Palomino, P., Falcao, T. P., Medeiros, R., Uehara, M., Bittencourt, I., and Mello, R. F. (2022). Plataformas de dados educacionais: Análise com foco no plano nacional de educação. In *Anais do I Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil*, pages 60–68. SBC.

- Portela, A. M., de Souza Leitão, G., da Silva Barreto, R., and de Oliveira, E. H. T. (2020). Uso de learning analytics sobre dados de simulados para apoio à avaliação da aprendizagem por professores e gestores. *RENOTE*, 18(2):388–397.
- Raj, A., Bosch, J., Olsson, H. H., and Wang, T. J. (2020). Modelling data pipelines. In *2020 46th Euromicro conference on software engineering and advanced applications (SEAA)*, pages 13–20. IEEE.
- 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.
- Schneider, T. F., Cechinel, C., de Mello, R. F., Freitas, E. L., da Rocha Falcão, T. P., and Ramos, V. F. (2022). Análise do nível de maturidade na adoção de learning analytics em instituições de ensino superior brasileiras. In *Anais do I Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil*, pages 11–20. SBC.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10):1380–1400.
- Viana, F. S., Santana, A. M., and Rabêlo, R. d. A. L. (2022). Avaliação de classificadores para previsão de evasão no ensino superior utilizando janela semestral. In *Anais do XXXIII Simpósio Brasileiro de Informática na Educação*, pages 908–919. SBC.