

Detecção de desistência de estudantes em disciplinas ofertadas com apoio do ambiente Moodle: uma discussão sobre resultados alcançados

Bernardo Dalfovo de Souza¹, Benjamin Grando Moreira¹, Carolina Bittencourt Wang¹

¹Universidade Federal de Santa Catarina (UFSC) – campus Joinville
Joinville – SC – Brasil

{bernardo.dalfovo, carolina.wang.ufsc}@gmail.com, benjamin.grando@ufsc.br

Abstract. *Education dropout is a problem that leads to significant social, economic, and personal consequences for students. In the context of higher education, dropping out of courses during undergraduate studies can be one of the factors contributing to overall dropout rates. This paper presents the development of indicators based on attendance and the completion of evaluative activities, extracted from the Moodle learning environment. The main result of this work is the introduction of indicators that can identify student dropouts with a recall of 87%, using 25% of the course duration. Additionally, this paper discusses related research, the metrics used to evaluate results, and the outcomes achieved.*

Resumo. *A evasão escolar é um problema que acarreta consequências sociais, econômicas e pessoais significativas para os alunos. No contexto do ensino superior, a desistência de disciplinas durante a graduação pode ser uma das causas que levam à evasão. Neste trabalho, apresentamos o desenvolvimento de indicadores com base na presença dos alunos e na realização de atividades avaliativas, extraídos do ambiente de aprendizagem Moodle. Como resultado principal deste estudo, são apresentados indicadores capazes de identificar a desistência dos alunos com uma sensibilidade de 87%, utilizando 25% do tempo total da disciplina. Além disso, são discutidos trabalhos semelhantes, as métricas utilizadas para avaliar os resultados e os resultados alcançados.*

1. Introdução

Durante a realização de um curso de graduação, percebe-se que diversos alunos não conseguem completar o cronograma de determinadas disciplinas por vários motivos, como falta de motivação e dificuldade em compreender o conteúdo. A não conclusão da matéria, que configura um tipo de fracasso escolar, a desistência, configura um prejuízo pessoal e profissional para o aluno, assim como de resultado sobre a produtividade das universidades e sociedade [Nagai and Cardoso 2017].

De acordo com o Ministério da Educação (MEC) [Ministério da Educação 1996], o conceito de evasão pode ser relativo ao curso (quando o discente se desliga apenas do curso superior em que estava matriculado) ou à instituição (o aluno deixa de frequentar a instituição), e ao sistema (o estudante deixa de frequentar o ensino superior). Baseando-se em literatura da área, o conceito de evasão em cursos de graduação é utilizado para identificar a evasão em cursos ou em disciplinas específicas, buscando prever antecipadamente a evasão de discentes [Manhães et al. 2011, Morais 2018, Queiroga et al. 2018].

O presente trabalho avalia especificamente a não-conclusão de disciplinas de um curso de graduação em modalidade presencial, tratando desse conceito como desistência ou abandono. Com o propósito de discernir indicadores que possam prenunciar o potencial abandono de uma disciplina acadêmica, foi realizado um estudo de caso. Os dados pertinentes para a análise foram extraídos dos registros de atividade no ambiente virtual de ensino Moodle (Modular Object-Oriented Dynamic Learning Environment). Importante salientar que a coleta desses dados foi efetuada diretamente no ambiente Moodle, pois várias instituições acadêmicas optam por evitar a instalação de complementos (plugins) em suas instâncias do sistema. O trabalho discute sobre as melhores formas de avaliar (aplicação de métricas de avaliação) para esse tipo de problema e apresenta os seguintes resultados diferentes em relação ao que é convencional outros trabalhos similares: (1) Avalia a desistência em disciplina e não a evasão de curso ou reprovação em disciplina; (2) Utiliza apenas dados da disciplina sendo analisada; e (3) Não utiliza técnicas de aprendizado de máquina para obter a classificação.

2. Fundamentação teórica

De acordo com [Pinto 2021], somente no ano de 2019, a quantidade de alunos que evadiram de cursos de graduação em universidades públicas chegou a 203.784, resultando em um custo contábil de R\$ 8.974.795.155,90 para as instituições públicas de ensino. Além do prejuízo contábil causado pela evasão nas universidades, o discente também arca com um custo econômico, muitas vezes sem perceber. Ao levar em consideração o valor do salário mínimo no ano avaliado, a autora descreve o custo econômico como os potenciais salários renunciados pelos alunos ao evadirem, o que resulta em uma média brasileira de R\$ 21.277,59 para cada evadido, representando cerca de um ano de trabalho.

A aplicação da análise de dados, especificamente na educação, é denominada Ciência de Dados Educacionais (EDS, na sigla em inglês para Educational Data Science) e combina habilidades técnicas e sociais à compreensão da prática educacional em diferentes ambientes de aprendizagem [Filatro 2020]. Ainda segundo [Filatro 2020], a EDS tem sua origem na década de 2000, nas conferências sobre Mineração de Dados Educacionais (EDM, da sigla em inglês para Educational Data Mining), e posteriormente, na conferência sobre Analítica da Aprendizagem (LA, sigla em inglês para Learning Analytics), sendo consolidada com pesquisas sobre a Inteligência Artificial na Educação (AIED, em inglês para Artificial Intelligence in Education).

Ao analisar dados, diversos modelos e técnicas podem ser empregados, dividindo as formas de análise em análise exploratória, explícita e implícita. Os resultados apresentados neste trabalho são obtidos a partir de uma análise explícita de dados.

Na análise explícita, a informação está disponível explicitamente nos dados, sendo necessária alguma operação para ressaltar o dado e produzir a informação. Diferentemente de uma análise exploratória dos dados, cujo objetivo é entender os dados, na análise explícita o objetivo é claro e específico [Amaral 2016]. Entre os procedimentos e técnicas para auxiliar na compreensão de dados, a EDS faz uso de modelos matemáticos construídos para comunicar insights a educadores, alunos, gestores, designers instrucionais e outras partes interessadas [Filatro 2020].

Entre os elementos da análise explícita estão¹, segundo [Amaral 2016]: a

¹Aqui são relacionados apenas aspectos da análise explícita de dados utilizados neste trabalho.

elaboração de predicados, que são condições lógicas para produzir um subconjunto de dados; a identificação de distorções, valores afastados dos demais, cuja distorção pode prejudicar os resultados; e a estratificação, cujos estratos são resumos de intervalos de uma variável numérica, definidos na geração do estrato.

Na análise de dados implícita, os objetivos mais comuns de análise envolvem a classificação, regressão, agrupamento e associação, frequentemente utilizando técnicas de aprendizado de máquina, tais como os algoritmos de árvores de decisão e Redes Neurais Artificiais, como mencionado em [Manhães et al. 2011, Burgos et al. 2018].

2.1. Trabalhos similares

Nesta seção, serão apresentados trabalhos que trazem uma aplicação similar à proposta deste estudo, embora esses trabalhos realizem uma análise implícita dos dados. Foram selecionados trabalhos nacionais e internacionais para discussões que abordam os objetivos deste estudo, incluindo os resultados obtidos, técnicas utilizadas, métricas de avaliação dos resultados e o grau de antecipação alcançado por meio da predição.

[Gottardo et al. 2014] utilizam o ambiente Moodle para obtenção dos dados e dividem as interações com o ambiente virtual em três categorias: 1) entre estudante e conteúdo; 2) entre estudante e professor; e 3) entre estudantes, supervisionados ou não. O objetivo do trabalho é prever o desempenho do estudante (a evasão é desconsiderada). Os alunos são divididos em intervalos de notas: maior que 88; entre 77 e 88; e menor que 77. A análise implícita realizada pelo trabalho utiliza os algoritmos *Random Forest* (RF) e *Multilayer Perceptron* (MLP). A avaliação dos resultados é feita utilizando uma matriz de confusão aplicada em três experimentos. O classificador RF apresenta acurácia média de 77,4%, 77,2% e 72,7% para cada um dos três experimentos, enquanto que o MLP resulta em acurácia média de 80,1%, 77,2% e 76,9%. Os autores não mencionam tentativas de utilização dos modelos com uma base de dados temporalmente reduzida, o que possibilitaria antecipar o resultado do desempenho acadêmico ao longo da disciplina, em vez de obter essas classificações apenas ao final do semestre.

Em [Queiroga et al. 2018], o objetivo envolve a evasão em um curso de graduação. O trabalho utiliza a contagem de interações dos estudantes no Moodle como a principal fonte de informação para a geração e comparação de modelos analíticos. Os algoritmos de aprendizado de máquina utilizados incluem *Bayes Net*, *Simple Logistic*, *MLP*, *RF* e *J48*. O trabalho também utiliza uma matriz de confusão como métrica de avaliação dos resultados. Um dos resultados mostra que o algoritmo *Simple Logistic* não alcança uma taxa de Verdadeiros Positivos (VP - alunos indicados como evadidos e que realmente evadiram) de 75% desde a primeira semana, mas gradualmente aumenta a taxa de acerto para 87% até a semana 25 (final do primeiro semestre) e se aproxima de 94% antes do final do segundo semestre da graduação. Para um segundo cenário, os cinco modelos ultrapassam a marca de 77% para VP desde a primeira semana, atingindo 99% a partir da quinta semana com o modelo RF.

Em [Burgos et al. 2018], cada atividade realizada pelos discentes ao longo de um semestre é utilizada. Para a construção dos modelos, é adotada uma metodologia incremental: o primeiro modelo utiliza apenas dados da atividade um; o segundo modelo incorpora dados das atividades um e dois; o terceiro modelo inclui as informações das notas das atividades um, dois e três; e assim por diante, até o modelo doze, em que todos

os dados são utilizados. Os autores classificaram os alunos entre desistentes e não desistentes, além de acrescentarem a informação da semana em que esses alunos evadiram o curso.

Foram aplicadas as seguintes técnicas para classificação: Rede Neural *feed-forward* (FFNN), Máquina de Vetores de Suporte (SVM), *Probabilistic Ensemble Simplified Fuzzy* e *Adaptive Resonance Theory Mapping*. Os resultados alcançados pelo modelo proposto indicam que, em média, o classificador proposto é capaz de detectar a desistência de um aluno até 1,6 semana antes da desistência. Isso é feito com uma precisão de 98,95%, sensibilidade de 96,73%, especificidade de 97,14% e acurácia de 97,13% na semana 10, que é, em média, quando ocorre a desistência nas turmas analisadas.

O trabalho de [Viana et al. 2022] avalia a evasão em relação aos cursos de graduação. A coleta de dados foi realizada por meio de atributos sociais, como raça, sexo, estado civil, idade e se o aluno foi selecionado por meio de um programa de cotas, bem como atributos acadêmicos, incluindo notas, quantidade de reprovações por nota e falta, e quantidade de aprovações. Os atributos acadêmicos coletados foram analisados em 6 conjuntos, representando semestres de forma cumulativa para alguns dos atributos (o primeiro conjunto possui apenas informações do primeiro semestre, enquanto o segundo conjunto possui informações do primeiro e segundo semestre). Os algoritmos de classificação utilizados incluíram Árvores de Decisão, RF, MLP, SVM, *K-Nearest Neighbors* (KNN) e *Gaussian Naive Bayes*. Cada algoritmo foi validado utilizando métricas como acurácia, precisão, sensibilidade, pontuação F1, índice Kappa e área sob a curva ROC. O melhor resultado foi obtido com o uso do RF, apresentando valores médios para os seis semestres de acurácia de 91,55%, sensibilidade de 0,92, precisão de 0,95, pontuação F1 de 0,93, índice Kappa de 0,81 e área sob a curva ROC de 0,91.

O trabalho de [da Silva Garcia et al. 2022] calcula a probabilidade de um discente ser reprovado ou aprovado em sua disciplina no início do semestre. Informações acadêmicas e pessoais dos alunos são extraídas a partir dos componentes curriculares coletados, como a fase letiva em que o aluno está cursando o componente, o resultado obtido (aprovação ou reprovação), sexo, data de nascimento e, por fim, todas as disciplinas que foram cursadas em fases anteriores ao semestre de análise. Os algoritmos utilizados para a análise incluíram *Naive Bayes*, IBK, JRIP, J48, RF e MLP. Os autores utilizam a acurácia e a sensibilidade como métricas principais para a avaliação dos resultados.

2.1.1. Avaliação dos trabalhos relacionados

A partir dos trabalhos similares é possível perceber que os problemas tratados são, em sua maioria, relacionados à evasão de curso e não à desistência de disciplina. Normalmente, leva-se em consideração o comportamento em vários semestres, bem como dados além daqueles obtidos em disciplinas, como a presença e notas.

A maioria dos trabalhos avaliou os resultados não considerando apenas a acurácia, mas também a precisão e sensibilidade. O uso da precisão e sensibilidade é apropriado, uma vez que a evasão/desistência são situações que deveriam ser atípicas, e a acurácia poderia mascarar os resultados obtidos. Por exemplo, considerando uma situação hipotética em que a desistência de uma turma seja de 20%. Uma regra que classifique todos os

alunos como não desistentes obteriam uma acurácia de 80% nas classificações, mas obteriam uma precisão/sensibilidade de zero.

Além das métricas de avaliação, é importante considerar a identificação dos alunos desistentes de forma antecipada. Ou seja, a classificação não é útil se considerar todo o desempenho do estudante no semestre, sendo necessário identificar o quanto antes para agir a fim de evitar a desistência. Nesse sentido, apenas os trabalhos [Queiroga et al. 2018, Burgos et al. 2018, da Silva Garcia et al. 2022] consideraram suas classificações levando em consideração essa antecipação.

3. Proposta do trabalho para classificação de desistência

Para que a fase de análise de dados possa tomar forma, torna-se necessária uma etapa de classificação dos dados, na qual decide-se quais características definem um determinado comportamento. No caso deste trabalho, o comportamento que se deseja identificar é a desistência.

Sobre as disciplinas analisadas, inicialmente são apresentadas análises de uma turma localizada no meio de cursos de engenharia. Essa disciplina é compartilhada entre três cursos diferentes. A segunda disciplina pertence à segunda fase e é ministrada para alunos de quatro cursos de engenharia. Ambas as disciplinas têm uma carga horária de 60 horas (4 créditos) e são ministradas em dois encontros semanais. Para identificar os alunos considerados desistentes no final do semestre de aplicação da disciplina, um professor da disciplina realizou uma marcação manual. Além disso, em ambas as disciplinas, foram aplicadas avaliações contínuas, incluindo questionários, atividades e fóruns, durante o semestre, quase semanalmente.

As análises partiram de dois relatórios básicos disponíveis no ambiente Moodle: o relatório de presença e o relatório de notas. Para obter esses relatórios, é necessário acessar o módulo de presenças do Moodle e, em seguida, selecionar a opção de exportação, que permite exportar os dados de presença no formato de planilha eletrônica padrão, como Microsoft Excel e OpenOffice. Para o relatório de notas, é preciso acessar o módulo de notas e, na sequência, selecionar a aba de exportação, também possibilitando a exportação dos dados no formato Microsoft Excel ou OpenOffice.

3.1. Obtenção dos dados - presenças e notas

O Moodle possui um plugin que permite realizar o controle de presenças² no ambiente. Esse recurso permite configurar aspectos diferentes na atribuição de presenças, alterando a quantidade de presença atribuída por sessão registrada. De maneira geral, o controle considera o valor máximo indicado pelas opções como uma presença completa, e graduações até a ausência, podendo nomear cada um desses indicativos. No arquivo de exportação, cada opção é apresentada por uma fração com o denominador sendo a maior pontuação indicada.

Cada atividade de avaliação presente no Moodle desempenha um papel fundamental na elaboração do relatório de desempenho acadêmico. Este relatório engloba avaliações individuais, categorias de avaliações e a nota global do curso, proporcionando uma visão abrangente do progresso do aluno. Importante salientar que, ocasionalmente, podem existir atividades registradas que, embora estejam no sistema, não tenham

²Link para o plugin de presenças do Moodle: https://moodle.org/plugins/mod_attendance

sido disponibilizadas ou avaliadas. É relevante observar que o relatório não detalha as ponderações associadas a cada atividade na composição final da nota do curso, tampouco fornece informações sobre as datas de realização das atividades ou se o aluno completou uma atividade que ainda não foi avaliada pelo professor.

3.2. Proposta de indicadores baseados em presenças e notas

Este trabalho propõe um sistema de pontuação e ranking para classificar os alunos como desistentes ou não desistentes. Para categorizar os alunos em quatro grupos, utilizou-se uma divisão em quartis com base na pontuação dos alunos. Essa pontuação varia de zero a 100, e cada uma das três categorias de dados (presença, notas e atividades importantes) contribui com 1/3 da pontuação total. Dessa forma, os alunos não são avaliados apenas em uma habilidade, mas sim em várias competências.

3.2.1. Pontuação de presença

A pontuação de presença é calculada a partir de pesos, atribuídos para cada aula e calculados a partir da Equação 1.

$$W_i = \frac{Me_i}{Max_i} \quad (1)$$

Para um dia i , Max_i representa a pontuação máxima definida pelo professor no Moodle, Me_i representa a média das pontuações do Moodle da classe para aquele dia e W_i é o peso atribuído a ele. Caso a presença não tenha sido registrada para o dia, Me_i é definido como a média de presenças da turma para o período analisado como um todo.

Para cada dia em que a presença foi computada para a turma, a pontuação é calculada de acordo com a Equação 2.

$$P_f = \sum_{i=1}^d \frac{Po_i \cdot W_i}{Max_i} \quad (2)$$

Nessa equação, P_f representa a pontuação final, Po_i é a pontuação atribuída pelo Moodle, W_i representa o peso calculado, Max_i é a pontuação máxima atribuída pelo Moodle, i é o dia de presenças analisado e d é o número total de dias no relatório de presença. Caso Po_i seja equivalente a Max_i , o peso não é levado em consideração.

3.2.2. Pontuação de conclusão de atividades

Para o cálculo de pontuação para a categoria de atividades concluídas, utilizou-se a Equação 3.

$$P'_f = \frac{Po_{max} - \sum_{j=1}^n (W'_j \cdot t_j)}{Po_{max}} \cdot 100 \quad (3)$$

Na qual P'_f representa a pontuação de atividades concluídas final, Po_{max} é a pontuação máxima a ser obtida através de todas as atividades, W'_j o peso definido para a

atividade j , t_j a taxa de conclusão da atividade j , com relação à turma, e n é o número total de atividades realizadas. Se a atividade j foi concluída, W'_j equivale a zero; caso contrário, é igual a um.

3.2.3. Pontuação de notas

A definição da pontuação para a categoria de notas é representada pela Equação 4. A equação realiza o cálculo do peso de uma nota de uma atividade ao relacionar a maior nota atingida na turma com a nota atingida pelo aluno analisado e então multiplica este valor pela média de notas da turma para a mesma atividade. O cálculo é realizado para todas as atividades, que são então somadas. O valor encontrado pela soma é subtraído e dividido da pontuação máxima atingível.

$$P''_f = \frac{(G_{max} - \sum_{k=1}^n (1 - \frac{G_k}{A_k}) \cdot M_k)}{G_{max}} \cdot 100 \quad (4)$$

Nessa equação, P''_f representa a pontuação final para a categoria de notas, G_{max} representa a maior soma de notas possível, G_i equivale à nota individual para a atividade i , A_i simboliza a maior nota obtida entre a classe para a atividade i , M_i é a média de notas da classe para a atividade i , e n é a quantidade total de atividades realizadas.

4. Aplicação dos indicadores elaborados na análise de turmas

Com base na pontuação gerada pelos indicadores propostos, os alunos foram divididos em quartis, sendo que aqueles nos dois quartis inferiores foram considerados como potenciais desistentes. A partir dessa classificação, o algoritmo foi avaliado com 25%, 50% e 75% dos dados disponíveis, visando entender o quão antecipadamente é possível prever o comportamento de desistência.

Depois de atribuir valores a cada aluno, os resultados foram ordenados e foi necessário determinar a partir de qual posição no ranking um aluno deveria ser considerado como potencial desistente. A recomendação é utilizar a taxa histórica de desistência na disciplina como referência e aplicar esse critério para realizar essa classificação. Para os resultados de análise das turmas apresentados a seguir, foi escolhido arbitrariamente o limiar de 50% (dois quartis inferiores) como indicativo de potencial desistência.

A primeira turma analisada, que será referida como "Turma 1", apresenta 41 alunos e 25 atividades realizadas. As matrizes de confusão mostradas na Figura 1 apresentam os resultados para 75%, 50% e 25% da aplicação da disciplina.

	Real	Desistente	Não desistente
Previsto			
Desistente		9	11
Não desistente		1	20

a) 75% da disciplina

	Real	Desistente	Não desistente
Previsto			
Desistente		9	11
Não desistente		1	20

a) 50% da disciplina

	Real	Desistente	Não desistente
Previsto			
Desistente		10	10
Não desistente		0	21

a) 25% da disciplina

Figura 1. Matriz de confusão da Turma 1: a) 75% do período; b) 50% do período; e c) 25% do período

Fonte: os autores (2023)

Na segunda turma analisada, que será referida como "Turma 2", havia 86 alunos matriculados e 23 atividades realizadas. As matrizes de confusão mostradas na Figura 2 apresentam os resultados para 75%, 50% e 25% da duração da disciplina.

Previsto \ Real	Desistente	Não desistente
Desistente	32	11
Não desistente	0	43

a) 75% da disciplina

Previsto \ Real	Desistente	Não desistente
Desistente	32	11
Não desistente	0	43

a) 50% da disciplina

Previsto \ Real	Desistente	Não desistente
Desistente	31	12
Não desistente	1	42

a) 25% da disciplina

Figura 2. Matriz de confusão da Turma 2: a) 75% do período; b) 50% do período; e c) 25% do período

Fonte: os autores (2023)

A acurácia do modelo é comparável com o de outros trabalhos, como [Martins et al. 2017, Gottardo et al. 2014], mas obtido de forma mais simples. As métricas foram avaliadas com três diferentes volumes de dados em duas turmas distintas, de forma a avaliar o que seria a utilização em uma aplicação prática: prever a desistência em tempo hábil para realizar alguma intervenção. Com 100%, 75% e 50% dos dados, a Turma 1 apresentou sensibilidade e precisão de 90% e 45%, respectivamente, e sensibilidade de 100% e precisão de 50% com 25% do total de dados. Os resultados para a Turma 2, por sua vez, foram de 100% para a sensibilidade e 74.42% para a precisão com 100%, 75% e 50% dos dados, enquanto que com 25% dos dados a sensibilidade e precisão atingiram 96.875% e 72.09%, respectivamente.

Os resultados apresentados são diretamente impactados pela escolha da quantidade de desistentes da turma considerada. Como o número real de desistentes da Turma 1 são de aproximadamente 24% e na Turma 2 a quantidade de desistentes foi de aproximadamente 37%, a Turma 1 teve pior precisão do que a Turma 2. Sendo assim, uma vez que um objetivo do trabalho era apenas utilizar dados isolados da turma analisada, esse objetivo não pode ser considerado alcançado por depender de informação passada da turma para melhor desempenho do resultado.

5. Conclusões

O presente artigo apresentou as etapas de análise explícita e propostas de avaliação dos dados para identificar a desistência de alunos em disciplinas. Diferente de outros trabalhos similares, que utilizam análise implícita com técnicas de aprendizagem de máquina, a proposta deste trabalho elabora índices a partir de abordagem explícita dos dados. O trabalho também se baseou nos relatórios mais simples de serem obtidos do ambiente Moodle (relatório de presença e relatório de notas), que permite uma utilização simplificada, embora com uma indicação incorreta elevada de não desistentes, o que acarreta desperdício de esforço com alunos que não iriam realmente desistir. Esses erros na classificação podem ser reduzidos a partir de uma observação histórica nas taxas de desistências das turmas.

Embora o projeto objetive o desenvolvimento de uma ferramenta para classificar alunos quanto a desistência, essa ferramenta está em elaboração. Um questionamento sobre ela é a necessidade dos dados do Moodle serem exportados e carregados na ferramenta, o que poderia ser questionado quanto a proteção desses dados. Para isso, mais do que ter um serviço disponibilizado para o público geral, a ferramenta deveria ter sua disponibilização pela instituição da disciplina em análise, ou simplesmente utilizar as equações propostas para basear os indicadores de ranking.

Referências

- Amaral, F. (2016). *Introdução à ciência de dados: mineração de dados e big data*. Alta Books Editora.
- Burgos, C., Campanario, M. L., la Peña, D., Lara, J. A., Lizcano, D., e Martínez, M. A. (2018). Data mining for modeling students' performance: a tutoring action plan to prevent academic dropout. *Computers & Electrical Engineering*, 66:541–556.
- da Silva Garcia, L. M. L., Lara, D. F., Gomes, R. S., e Cazella, S. C. (2022). Mineração de dados educacionais na predição do desempenho acadêmico: um prognóstico a partir do percurso curricular realizado. In *Anais [...]*, Manaus. XXXIII Simpósio Brasileiro de Informática na Educação.
- Filatro, A. C. (2020). *Data Science na Educação: Presencial, a Distância e Corporativa: Presencial, a Distância e Corporativa*. Saraiva Educação SA.
- Gottardo, E., Kaestner, C. A. A., e Noronha, R. V. (2014). Estimativa de desempenho acadêmico de estudantes: análise da aplicação de técnicas de mineração de dados em cursos a distância. *Revista Brasileira de Informática na Educação*, 22(1):45–56.
- Manhães, L. M. B., Cruz, S. M. S., Costa, R. J. M., Zavaleta, J., e Zimbrão, G. (2011). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In: XXII Simpósio Brasileiro de Informática na Educação.
- Martins, L. C. B., Carvalho, R. N., Carvalho, R. S., Victorino, M. C., e Holanda, M. (2017). Early prediction of college attrition using data mining. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1075–1078. IEEE.
- Ministério da Educação (1996). *Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas*. BRASIL. Ministério da Educação, Brasília.
- Morais, A. M. (2018). *Abordagem avaliativa multidimensional para previsão da evasão do discente em cursos online*. Tese (doutorado em ciência da computação), Centro de Engenharia Elétrica e Informática, Universidade Federal de Campina Grande.
- Nagai, N. P. e Cardoso, A. L. J. (2017). A evasão universitária: uma análise além dos números. *Revista Estudo & Debate*, 24(1):193–215.
- Pinto, S. C. (2021). *Os custos da evasão de discentes das universidades brasileiras na modalidade de ensino presencial: uma perspectiva de custos contábeis e custos econômicos*. Dissertação (mestrado em ciências contábeis), Unidade Acadêmica de Pesquisa e Pós-Graduação, Universidade do Vale do Rio dos Sinos.
- Queiroga, E. M., Lopes, J. L., Araujo, R. M., e Cechinel, C. (2018). Modelo de predição da evasão de estudantes em cursos técnicos a distância a partir da contagem de interações. *Revista Thema*, 15(2):425–438.
- Viana, F. S., Santana, A. M., e de Andrade Lira Rabêlo, R. (2022). Avaliação de classificadores para predição de evasão no ensino superior utilizando janela semestral. In *Anais [...]*, Manaus. XXXIII Simpósio Brasileiro de Informática na Educação.