

Análise de Interações do Ambiente Virtual de Aprendizagem para Predição de Evasão em Cursos no Ensino a Distância

Interaction Analysis of the Virtual Learning Environment to Predict Evasion in Distance Learning Courses

Thiago Siqueira Sonnenstrahl^{1*}

Giliane Bernardi¹

Solange Pertile¹

¹Universidade Federal de Santa Maria - UFSM - Av. Roraima, 1000, Bairro Camobi
Santa Maria - RS - Brasil - *thiago.sonnenstrahl@iffarroupilha.edu.br

Resumo. Este artigo tem como objetivo, através da Mineração de Dados Educacionais (MDE), analisar, por meio da interação dos alunos no Ambiente Virtual de Aprendizagem (AVA), possíveis evasões em cursos na modalidade a distância, disponibilizando dados estratégicos para a tomada de decisão pelos gestores educacionais da instituição. Para a realização dos experimentos foram utilizados dois conjuntos de dados contendo as interações dos alunos no AVA Moodle de duas turmas subsequentes na modalidade EaD. Como resultado, a taxa de acerto (número de alunos que os algoritmos classificaram corretamente como evadido ou concluintes) do primeiro conjunto de dados foi de 93%, obtida com o algoritmo *Random Forest*, já para o segundo conjunto, a taxa de acerto foi de 85% com o modelo gerado pelo algoritmo *J48*. Os resultados da mineração apresentaram a tarefa_visualizada e o material_visualizado como indicadores de evasão. Desta forma, pode-se constatar que os alunos que pouco interagem com os recursos dentro do ambiente e não visualizam os materiais e tarefas disponibilizadas são mais propensos a evadirem do curso.

Palavras-chave: Educação a distância. Evasão. Mineração de dados educacionais.

Abstract. This article aims, through Educational Data Mining (MDE), to analyze, through the interaction of students in the Virtual Learning Environment (AVA), possible dropouts in distance learning courses, providing strategic data for decision making by the institution's educational managers. In order to carry out the experiments, two sets of data were used containing the interactions of students in

the AVA Moodle from two subsequent classes in distance learning mode. As a result, the hit rate for the first set of data was 93%, obtained with the Random Forest algorithm, while for the second set, the hit rate was 85% with the model generated by the J48 algorithm. The mining results dispelled the visualized_task and the visualized_material as evasion indicators. Thus, it can be seen that students who have little interaction with the resources within the environment and do not see the materials and tasks available are more likely to drop out of the course.

Keywords: Distance education. Evasion. Educational data mining.

1. Introdução

A modalidade de ensino a distância (EaD) tem apresentado todos os anos um crescimento em diferentes níveis de ensino no Brasil. Esse destaque na educação fornece-nos dados que comprovam o tamanho da expansão, como por exemplo, o aumento do número de instituições que contam com polos presenciais, mesmo com a flexibilidade da oferta desses polos ter deixado de ser obrigatória pelo Ministério da Educação - MEC (CENSO EAD.BR 2018, 2019).

A maior expansão dos polos em 2018 deu-se nas instituições privadas, com a criação de 2.577 polos nas com fins lucrativos e de 386 polos nas sem fins lucrativos. Em 2017, os números foram de 1.476 e 531, respectivamente. Entretanto, também houve uma expansão na criação de polos nas instituições públicas estaduais e federais, 276 e 102 em 2018, e 157 e 81, em 2017, respectivamente (CENSO EAD.BR 2018, 2019). Além disso, o número de cursos ofertados totalmente a distância teve um aumento significativo em 2018, chegando a um crescimento de 12.180 em relação a 2017.

Com esse avanço da oferta de polos e cursos EaD, o número de alunos também cresceu consideravelmente, sendo contabilizado um total de 9.374.647 alunos em 2018 (CENSO EAD.BR 2018, 2019). Tais dados mostram a capacidade e o crescimento da modalidade EaD em diversas áreas do Brasil, potencializando ainda mais a expansão da educação nessa modalidade.

Nesse contexto, surge também a evasão, um tema que tem preocupado os gestores das instituições e exigido uma atenção especial deste no âmbito educacional. A evasão está presente em todas as modalidades de ensino, seja presencial, semipresencial ou a distância (BITTENCOURT; MERCADO, 2014).

O abandono do aluno representa um prejuízo para diferentes partes, seja para a instituição como para o próprio aluno. A evasão pode estar ligada a diversas causas, que estão diretamente relacionadas à qualidade da educação oferecida pela instituição de ensino frequentada pelo aluno, ao ambiente escolar, à relação familiar dele, ao meio social em que ele vive ou a motivos concernentes à vida pessoal do próprio aluno.

Nesse sentido, os motivos podem estar relacionados interna ou externamente à instituição, e, independentemente da categoria em que estejam inseridos, esses fatores precisam ser

tratados pela gestão ou pelo órgão ao qual o aluno esteja ligado. Ações ou práticas pedagógicas são necessárias para incentivar o aluno, de forma que ele não se desestimele ou venha a evadir, prejudicando assim o aumento do índice de efetividade na educação e, consequentemente, um dos objetivos das IES, a permanência e o êxito dos estudantes no curso.

Apesar de 88,2% das instituições informarem ao Censo de 2018 (CENSO EAD.BR 2018, 2019) conhecerem os motivos da evasão, diferente do apresentado ao Censo de 2017, em que 50% desconheciam esses motivos, considera-se que o aumento da oferta de cursos e o crescimento no número de matrículas possa elevar a probabilidade de evasão e o acompanhamento mais próximo dessas taxas pelas instituições.

Com o propósito de buscar medidas para auxiliar na redução da evasão, a mineração de dados surge como excelente alternativa, pois possibilita identificar um padrão de fatores de desistência do estudante, podendo através deste padrão prever os estudantes que possam vir a evadir, antes que a evasão ocorra. Nesta proposta, trabalharemos apenas com fatores internos dos estudantes, como por exemplo, as interações dos estudantes no AVA, uma vez que fatores externos são mais difíceis de se adquirir.

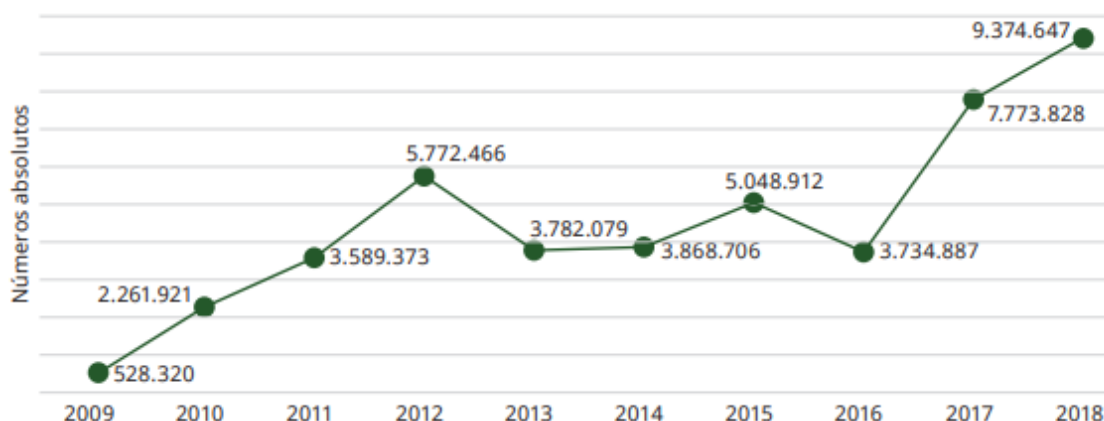
2. Matrículas versus Evasão no EAD

A EaD é uma modalidade de ensino em que alunos e professores encontram-se em espaços distintos. Segundo Moran (2002), pode-se definir que a educação a distância é o processo de ensino-aprendizagem mediado por tecnologias, no qual professores e alunos estão separados espacial e/ou temporalmente. Isto é, a EaD tem se mostrado bastante rica em potenciais pedagógicos e de democratização do conhecimento (MILL, 2013).

Nesse sentido, a tecnologia é fundamental para que a relação entre aluno e professor aconteça de forma dinâmica e eficiente, apropriando-se de diferentes recursos tecnológicos e inovadores, proporcionando aos seus alunos um ambiente capaz de buscar a evolução e o aprendizado nos diferentes espaços educacionais ofertados por essa modalidade de ensino. No âmbito do estudante, nota-se que o ensino a distância proporciona acesso à capacitação e formação de indivíduos que estão em locais distantes dos grandes centros de ensino, bem como permite uma maior flexibilidade e autonomia em relação ao estudo, sem perder a qualidade no ensino. Vale ressaltar que cabe ao aluno administrar seu próprio aprendizado e ter responsabilidades em estabelecer seus próprios objetivos quanto ao seu estudo (BELLONI, 2001).

De acordo com os dados apresentados no Figura 1, pode-se perceber um crescimento significativo no número de matrículas na educação a distância, o que indica a necessidade e a devida importância nessa modalidade para o crescimento da educação no Brasil.

Figura 1 – Número de matrículas na EaD entre 2009-2018



Fonte: CENSO EAD.BR 2018 (2019)

Pesquisas da Associação Brasileira de Educação a Distância (CENSO EAD.BR 2016, 2017) indicam os principais problemas enfrentados pela EaD no Brasil: o principal é a evasão, foco deste estudo, caracterizada pela desistência do estudante do curso; o segundo é a resistência dos docentes em relação à modalidade, muitas vezes causada justamente pela falta de conhecimento sobre TIC e sobre o devido domínio nas plataformas EaD. Por fim, a dificuldade de adaptação dos estudantes à modalidade EaD. Esses problemas podem causar certa frustração aos alunos devido à metodologia de ensino ser diferente da abordada na modalidade presencial, pois deixamos de ter o contato direto com o professor, acabando, até mesmo, dando equivocadamente menor importância ao ensino a distância.

A evasão, que se torna um dos maiores desafios enfrentados pela EaD, é definida, segundo Maia (2007), como a desistência do aluno em completar o curso, independentemente se cursou aulas ou não, ou seja, aquele que desiste definitivamente do curso em qualquer etapa do curso (FAVERO, 2006; ABBAD; CARVALHO; ZERBINI, 2006). Outros autores, como Toczec et al. (2004), a definem como o desligamento ou abandono do aluno da instituição de ensino, que pode ser compreendido como um processo individual, mas também pode constituir-se em coletivo.

Para Eyng et al. (2013), a evasão é definida como um processo de abandono da escola no ano letivo, levando o estudante a deixar de frequentar as aulas, bem como de realizar suas atividades curriculares, ocorrendo de forma gradativa no andamento do curso, até mesmo sem que isso seja notado pelos gestores escolares. Percebe-se que isso ocorre no decorrer das aulas do curso, sem ser observado de forma a evitar que o abandono do aluno ocorra, destacando ainda mais o papel primordial do professor para mitigar a evasão na educação.

Martínez, García e Montoro (2003) ainda classificam a evasão de quatro formas: *i*. Evasão: quando o aluno abandona o curso durante o seu período sem retornar; *ii*. Trancamento: interrupção temporária do curso; *iii*. Evasão do curso: o aluno abandona o curso antes do término, porém obtém conhecimento; *iv*. Não iniciado: o aluno não inicia o curso.

Já para Manhães et al. (2011), a evasão pode ser definida em três eixos: *i*. A evasão de curso: caracterizada por abandono do estudante, desistência, transferência ou, até mesmo,

pela sua exclusão do curso por norma institucional; *ii*. Evasão da instituição: o estudante desliga-se da instituição na qual está matriculado; *iii*. Evasão do sistema: abandono definitivo ou temporário.

A evasão pode estar ligada a diferentes motivos particulares de cada IES e também a cada necessidade não correspondida para as diferentes realidades dos alunos inseridos nesse campo educacional. De acordo com os censos anuais, as taxas de evasão são maiores nos cursos a distância quando comparadas com os cursos presenciais. De acordo com o Censo EAD.BR 2018 (2019), a maior porcentagem de instituições (22,2%) encontra-se nas taxas de evasão entre 26% e 50%, quanto a oferta de cursos totalmente a distância. Em 2017, esse número era de 6%.

Segundo Manhães et al. 2011, para reduzir o problema da evasão, é primordial a detecção dos alunos com tendência a evadir, de modo que possamos atentar especificamente ao aluno ou grupo de alunos, dispondo de atendimento diferenciado para tratar o problema. Deve-se, portanto, detectar os sinais enviados pelos alunos quando estão em situação de risco (DORE; LÜSCHER, 2011).

Como forma de buscar sempre a eficiência de uma instituição de ensino, os gestores cada vez mais buscam ferramentas e métodos que os auxiliem a mitigar a evasão do aluno, sendo que identificar indicadores de evasão e os motivos que levam o aluno a evadir do curso matriculado pode auxiliar os gestores na tomada de decisão (OLIVEIRA et al., 2017).

3. Mineração de Dados Educacionais

A mineração de dados trata-se de extrair ou minerar conhecimentos em grandes bases de dados. Esta é uma área cujo foco principal é desenvolver ou adaptar métodos e algoritmos para explorar um conjunto de dados produzidos e coletados em ambientes educacionais (BAKER; ISOTANI; CARVALHO, 2011). A mineração de dados educacionais (MDE) busca compreender os dados produzidos por alunos e professores, bem como suas interações por meio de diferentes ferramentas dentro de um AVA.

A MDE pode ser aplicada, por exemplo, para descobrir os motivos que influenciam o aprendizado do aluno, analisar o comportamento do aluno evadido, entre outras possibilidades. Autores (RABELO et al. 2017; BURGOS et al., 2017; QUEIROGA, CECHINEL e ARAÚJO, 2017; SCHMITT, 2018; GONÇALVES e BELTRAME, 2019) têm focado suas pesquisas no uso de mineração de dados para prever possíveis evasões de estudantes. A MD compreende a aplicação de técnicas e algoritmos sobre os dados em busca de conhecimento útil e implícito (GOLDSCHMIDT; BEZERRA; PASSOS, 2015). Segundo Romero e Ventura (2013), a maioria das técnicas tradicionais de mineração de dados, tais como classificação, agrupamento e técnicas de análise de associação, já foram aplicadas com êxito no domínio da educação. A seguir, serão descritas, de forma sucinta, as principais técnicas de MD.

1) Associação: a associação apresenta padrões de relacionamento entre itens de uma base de dados. Ela busca os fatos que tendem a ocorrer de forma simultânea e com frequência. Poderíamos minerar os dados com base nas notas dos alunos em suas disciplinas, por exemplo, “90% dos alunos que têm bom desempenho na disciplina de Lógica têm bom desempenho em programação”. São exemplos de algoritmos que implementam essa técnica: Apriori, GSP, DHP, entre outros (ZAKI, 2000).

2) Classificação: é o processo de encontrar um conjunto de modelos que descrevem e distinguem classes ou conceitos, com o propósito de utilizar o modelo para prever a classe de objetos que ainda não foram classificados, ou seja, ela busca prever uma classe de um novo dado automaticamente. Essa técnica é conhecida como predição, podendo ser de dois tipos: classificação ou estimação. Como exemplo, podemos utilizar um modelo de classificador para identificar quais são as principais causas da desistência de uma determinada disciplina. Com isso, podemos fazer conclusões, por exemplo, “que a desistência do aluno está relacionada com a faixa etária entre 25 e 30 anos”. Redes Neurais, Árvore de Decisão, Algoritmos Genéticos e Lógica Nebulosa são exemplos de algoritmos que podem ser aplicados na classificação (CASTRO; FERRARI, 2016).

3) Agrupamento: como o próprio nome já diz, é a técnica de identificar e aproximar registros similares, também conhecidos como clusterização. É a coleção de registros similares entre si. Classificar, estimar ou prever não é a pretensão do agrupamento, ele serve para identificar os grupos de dados similares. Com essa técnica, espera-se conhecer novos atributos alvos (rótulos) a partir de um conjunto de dados, sem ter classificação prévia (COSTA, et al., 2013). Como exemplo, podemos buscar a construção do modelo de análise e tratamento da evasão e retenção discente em um determinado curso, permitindo reconhecer tipos de classes distintas de alunos por meio de dados similares dos alunos. Para a realização dessa técnica, podem ser utilizados os algoritmos: *k-Means*, *k-Modes*, *kPrototypes*, Mapas de *Kohonen*, entre outros (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

3.1 Validação dos Algoritmos de Mineração de Dados

Para avaliar os resultados dos algoritmos foi utilizada a técnica de avaliação cruzada, do inglês *cross-validation*, também conhecida como validação cruzada em *k*-pastas, do inglês *k-fold cross-validation*. Ela consiste na divisão da base de dados em *k* subconjuntos, sendo *k*-1 conjuntos para treinamento dos algoritmos e o conjunto restante para teste. O processo consiste em dividir o conjunto de dados de forma aleatória, para a validação do algoritmo em *N* subconjuntos de dados igualmente distribuídos em teste e treinamento. Esse processo é realizado para todos os *k* subconjuntos, e a média do desempenho para os conjuntos analisados é utilizada como indicador de qualidade de desempenho (CASTRO; FERRARI, 2016). A validação cruzada adotando *k*=10 é a mais utilizada para essa finalidade, sendo também usada neste trabalho.

Para a comparação de desempenho, são empregadas as métricas de Acurácia, Precisão, Revocação e Medida-F. Para facilitar a compreensão dessas métricas, suas fórmulas serão demonstradas a seguir:

- **Acurácia (*accuracy*):** porcentagem de amostras positivas e negativas classificadas corretamente sobre a soma de amostras positivas e negativas, ou seja, é a taxa de sucesso do classificador. A acurácia é formalmente definida como a porcentagem dos exemplos de teste que são corretamente classificados (SHALEV-SHWARTZ; BEN-DAVID, 2014). Diante de um determinado conjunto de teste, é obtido o percentual dos exemplos que foram corretamente classificados. O cálculo de acurácia é dado por:

$$\text{Acurácia} = \frac{\text{Número de Classificações Corretas}}{\text{Número de Instâncias}}$$

Em que o Número de Classificações Corretas é a quantidade total de classificações previstas corretamente, tanto relacionadas aos alunos que foram classificados como desistentes quanto aos permanentes. Já o Número de Instâncias representa o número de alunos avaliados.

- **Precisão (*precision*):** determina o percentual de acertos do algoritmo em relação à previsão realizada. O cálculo é dado por:

$$\text{Precisão} = \frac{\text{Nro. Alunos Previstos} \cap \text{Nro. Alunos} \frac{\text{Evadidos}}{\text{Concluídos}}}{\text{Nro. Alunos Previstos}}$$

Em que o Número de Alunos Previstos representa a quantidade de alunos que foi classificada, no âmbito da pesquisa, como “Evadido” ou “Concluídos”, já que o objetivo é encontrar os alunos com tendência à evasão. O número de alunos evadidos compreende a quantidade de alunos que, de fato, desistiu.

- **Revocação (*recall*):** mede a integridade do algoritmo em relação aos alunos evadidos, ou seja, descreve a porção que foi classificada corretamente como exemplos positivos. A medida é dada pela seguinte equação:

$$\text{Revocação} = \frac{\text{Nro. Alunos Previstos} \cap \text{Nro. Alunos} \frac{\text{Evadidos}}{\text{Concluídos}}}{\text{Nro. Alunos} \frac{\text{Evadidos}}{\text{Concluídos}}}$$

Em que o Número de Alunos Previstos representa a quantidade de alunos que foi classificada, no âmbito da pesquisa, como “Evadido” ou “Concluídos”. Em suma, a

métrica de revocação é calculada pelo número de alunos previstos e evadidos/concluídos pelo número de alunos que evadiram/concluíram.

- Medida-F (*f-Measure*): métrica de avaliação que consiste na média harmônica de Precisão e Revocação e é definida pela equação:

$$Medida-F = \frac{(2 * Precisão * Revocação)}{(Precisão + revocação)}$$

4. Metodologia

Para o desenvolvimento deste trabalho foram utilizados dados de duas fontes distintas. No AVA Moodle foram extraídas as interações dos alunos, as quais consistem de LOG's gerados pelo sistema durante cada ação realizada pelo estudante no ambiente. Os dados extraídos e utilizados podem ser visualizados no quadro 1.

Quadro 1 – Atributos do AVA Moodle utilizados na MD

Tarefa submetida (<i>Assign_submitted</i>)
Tarefa visualizada (<i>Assign_viewed</i>)
Mensagem enviada (<i>Chat_sent</i>)
Mensagem visualizada (<i>Chat_viewed</i>)
Pasta visualizada (<i>Folder_viewed</i>)
Visualização no fórum (<i>Forum_viewed</i>)
Material visualizado (<i>Resource_viewed</i>)
Link visualizado (<i>url_viewed</i>)
Classe do aluno (evadido, concluído)

Fonte: Autor (2020).

A outra fonte de dados utilizadas foi o sistema de gestão da instituição, a qual contém as informações do aluno, tais como desempenho acadêmico, dados pessoais e informação sobre o *status* da sua matrícula (evadido ou concluinte).

Foram utilizados dados acadêmicos dos estudantes dos cursos subsequentes de Administração e Agroindústria ofertados na modalidade EaD. Os cursos analisados são ofertados em diversos polos de apoio presencial, possuindo disciplinas em sua estrutura curricular distribuída em três semestres letivos, conforme projeto pedagógico de cada curso. Os dados foram extraídos do início ao término da oferta de cada um dos cursos, tendo informações sobre os alunos que não concluíram o curso (classe de evadidos) e os alunos concluintes (classe de concluintes/formados).

Para a etapa de mineração de dados foi utilizada a ferramenta WEKA¹. A escolha dessa ferramenta justifica-se pelos diversos estudos e abordagens na literatura, além das suas funcionalidades e características, tais como a facilidade de utilização, o número de algoritmos disponíveis para a utilização e a possibilidade de alteração dos parâmetros de execução destes.

Após os dados serem selecionados e extraídos, resultando em mais de 14 mil linhas, foi realizada a preparação ou, como é conhecido, pré-processamento de dados, com os componentes e as ações que continham alguma interação. Essa seleção de componentes e ações deu-se por meio da visualização da contagem do número de interações em cada componente pelos alunos. Sendo assim, componentes e ações em que não existia interação alguma foram desconsiderados devido à não utilização do recurso durante os três semestres.

O próximo passo foi remover os dados dos alunos que não continham nenhuma interação durante todo o curso, ou seja, alunos sem qualquer tipo de interação dentro do ambiente foram removidos como forma de não prejudicar a predição pelos algoritmos. A justificativa para não haver interação, segundo a Diretoria de Educação a Distância, é devido a esses alunos nunca terem acessado o Moodle. Portanto, para este trabalho, foram considerados apenas os alunos que iniciaram o curso e apresentaram alguma interação no ambiente virtual.

Para realização dos experimentos, os dados foram divididos em três conjuntos:

- i) Conjunto 1: informações dos alunos do curso de Administração. Com um total de 180 instâncias (alunos), sendo 104 correspondentes a classe “concluído” e 76 da classe “evadido”. A taxa de evasão neste curso foi de 42%.
- ii) Conjunto 2: informações dos alunos do curso de Agroindústria. O conjunto contém um total de 161 instâncias (alunos), sendo 78 instâncias correspondentes a alunos concluintes e 83 de alunos evadidos. Neste curso, a taxa de evasão foi de 52%.

5. Resultados

Após a seleção dos atributos e preparação dos dados, ocorre a etapa de modelagem, por meio de testes e implementações dos algoritmos. A seleção dos algoritmos utilizados deu-se por meio dos trabalhos correlatos e de sua análise, bem como de sua disponibilidade na ferramenta WEKA. É importante salientar que para os experimentos são utilizados conjuntos de dados anotados para poder fazer a avaliação dos resultados.

Dessa forma, foram escolhidos os seguintes algoritmos para a realização dessa etapa: *J48*, *Naive Bayes*, *MultiLayer Perceptron*, *Randon Forest*, *IBK*, *SMO*, *OneR*. e para a geração dos modelos, foi utilizado o método *cross-validation*, assumindo um valor de 10 pastas.

¹ <https://www.cs.waikato.ac.nz/ml/weka/>

Ao analisarmos a Tabela 1, podemos perceber que os dados são semelhantes entre os algoritmos, ou seja, o modelo gerado por ambos os algoritmos acertou na maioria das vezes em que identificou que um aluno evadirá ou concluirá o curso, apresentando uma média de acurácia superior a 90%. O algoritmo *Naive Bayes* apresentou o menor índice de acertos para todas as métricas de avaliação, com taxa de 83,88% para a acurácia. Por fim, o algoritmo com melhor taxa foi o *Randon Forest*, apresentando 93,33%.

Tabela 1 – Métricas de desempenho dos algoritmos utilizados: Conjunto 1

Algoritmo	Métricas de avaliação			
	Acurácia	Precisão	Revocação	Medida-F
<i>J48</i>	91,66%	91,70%	91,70%	91,70%
<i>NaiveBayes</i>	83,88%	84,50%	83,90%	84,00%
<i>MultiLayerPerceptron</i>	90,00%	90,10%	90,00%	90,00%
<i>Randon Forest</i>	93,33%	93,33%	93,33%	93,33%
<i>IBK</i>	90,00%	90,00%	90,00%	90,00%
<i>SMO</i>	92,22%	92,30%	92,20%	92,20%
<i>OneR</i>	92,77%	93,10%	92,80%	92,70%

Fonte: Autor (2020).

A Tabela 2 apresenta o resultado dos algoritmos que foram aplicados no segundo experimento, tendo a maior taxa de acurácia obtida pelo algoritmo *J48*, com 85,71%. Assim como o primeiro experimento, os modelos gerados pelos algoritmos apresentaram resultados similares em ambas as métricas.

Tabela 2 – Métricas de desempenho dos algoritmos utilizados: Conjunto 2

Algoritmo	Métricas de avaliação			
	Acurácia	Precisão	Revocação	Medida-F
<i>J48</i>	85,71%	88,44%	85,70%	85,50%
<i>NaiveBayes</i>	79,50%	81,50%	79,50%	79,10%
<i>MultiLayerPerceptron</i>	84,47%	84,50%	84,50%	84,5%
<i>Randon Forest</i>	80,74%	80,80%	80,70%	80,70%
<i>IBK</i>	83,22%	83,20%	83,20%	83,20%
<i>SMO</i>	81,98%	83,80%	82,20%	81,70%
<i>OneR</i>	83,22%	84,90%	83,20%	83,10%

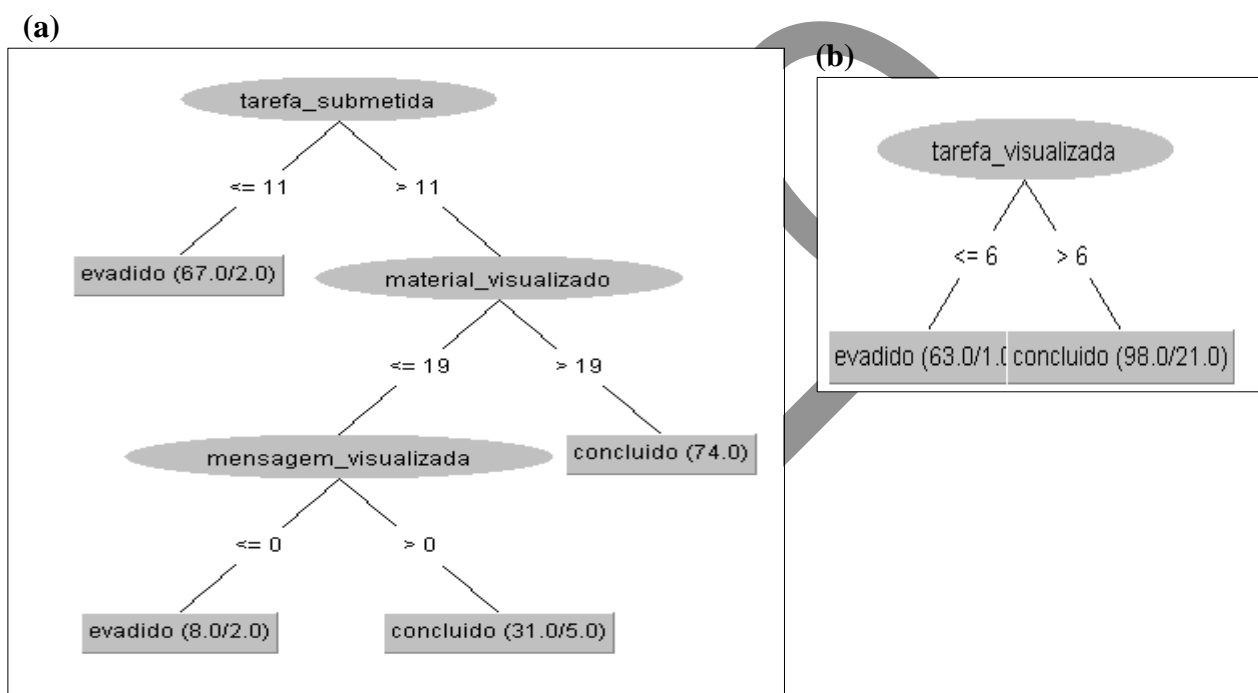
Fonte: Autor (2020).

O pequeno desvio apresentado nos resultados das métricas de avaliação em ambos os experimentos se dá pelo número de alunos que os modelos gerados retornaram como

sendo alunos evadidos/concluídos ser igual ao número de alunos evadidos/concluídos existentes nos conjuntos de dados.

Conforme pode-se observar na árvore de decisão gerada pelo algoritmo *J48* (Figura 2a), entre os atributos analisados, o atributo que determinou a evasão do aluno foi o recurso tarefa submetida. Ou seja, a árvore mostra que estudantes que enviam menos de 11% das tarefas (submissão de tarefas) são mais propensos a evadirem do curso. Já no segundo conjunto de dados (Figura 2b), o recurso que determinou a evasão pelo algoritmo foi tarefa visualizada. Esse fato pode ser devido a pouca exploração dos demais recursos e tarefas no AVA Moodle pelo curso analisado.

Figura 2 – Árvore de decisão gerada pelo *J48* para o (a) Conjunto 1 e (Conjunto 2)



5.1 Discussão

Ao analisar os resultados dos experimentos com o algoritmo *J48*, pode-se identificar que os modelos gerados apontaram para a submissão de tarefas e a visualização dos materiais como principais atributos para prever se um aluno está propenso a evadir ou não. Em relação aos resultados apresentados pelos algoritmos utilizados, apesar de haver algumas diferenças em alguns parâmetros e na implementação, ambos apresentaram resultados muito similares em ambos os conjuntos de dados, o que aumenta a confiabilidade na avaliação da saída dos mesmos.

Com base na análise das interações dos alunos com os recursos do AVA, pode-se observar que recursos como o *chat* – que é utilizado para a mediação pedagógica entre estudante e professor e requer participação síncrona, possui pouca adesão dos estudantes. Isso pode ser em razão da indisponibilidade de tempo no momento do *chat* ou também pelos alunos focarem mais na realização de atividades que envolvem processos avaliativos. O primeiro

conjunto de dados utilizado para os experimentos não apresentou nenhuma interação no recurso *chat* pelos alunos evadidos. Já o segundo conjunto de dados não possuía nenhuma interação neste recurso tanto pelos alunos evadidos como pelos concluintes, podendo neste caso, não ter sido utilizado pelos professores durante o curso.

Ainda, sobre a frequência nos fóruns e *chats*, destaca-se que a presença virtual do estudante é definidora no processo de ensino-aprendizagem, mas ela está relacionada à presença virtual e interação do docente e do tutor no sentido de acolher e provocar o estudante. Além disso, em determinados grupos de estudantes, as barreiras tecnológicas e de tempo para a dedicação à rotina de estudos faz com que esse estudante não participe de atividades síncronas, ou não busque esclarecer suas dúvidas, o que justifica a pouca utilização do chat e, até mesmo, do fórum.

Sendo assim, a qualificação e capacitação dos docentes e tutores é fundamental para a transposição didática dos conteúdos e construção de identidade no trabalho docente virtual. Somado a isso, conhecer a realidade do estudante, seu contexto social, sua fluência tecnológica e suas rotinas de trabalho e estudo auxilia o polo de apoio presencial e os tutores e professores a buscarem subsídios de inserção desses alunos no curso e na instituição. Essas questões podem ser inicialmente averiguadas a partir de extrações e análises de dados, como as deste estudo.

6. Conclusões

Neste trabalho, foi possível compreender a relevância do tema evasão no âmbito educacional, bem como as discussões sobre suas principais causas e abordagens, por diferentes autores na literatura. A permanência e o êxito dos estudantes são um tema importante para qualquer instituição de ensino e são do interesse dos gestores educacionais nos diferentes níveis em que atuam; por isso, observar fatores relacionados à evasão faz-se tão importante.

Na educação a distância, por meio do AVA Moodle, temos uma ferramenta que precisa ser explorada em seus mais diversos recursos, de forma que possamos utilizar melhor sua capacidade e, também, potencializar a mediação didática, qualificando a interação entre aluno, tutor, conteúdo e professor.

Deste modo, este trabalho buscou, por meio da mineração de dados educacionais e da interação do aluno no ambiente Moodle, dados e indicadores que pudessem ajudar os gestores a conhecer melhor a usabilidade da ferramenta, as características dos alunos com tendência a evadir, auxiliando na tomada de decisão pela instituição. Essas decisões envolvem qualificação da equipe multidisciplinar e acompanhamento da frequência virtual dos estudantes para além do acesso e observação dos fatores socioculturais e experienciais dos ingressantes. No entanto, foi verificada pouca utilização dos recursos do Moodle pelos tutores e professores, limitando a utilização de mais recursos na mineração.

Sendo assim, os dados da mineração mostram uma taxa de acerto para o primeiro conjunto de dados de 93%, obtido com o algoritmo *Random Forest*, e para o segundo conjunto, a taxa de acertos foi de 85% com o modelo gerado pelo algoritmo *J48*. Os melhores atributos que realizaram a predição foram a tarefa_visualizada e material_visualizado, ou seja, os indicadores de evasão para os cursos minerados foram a visualização de tarefas e de materiais disponibilizados. Isso leva à dedução do perfil do aluno concluinte nesses cursos, ficando este submetido a apenas materiais e tarefas disponibilizadas, buscando responder às tarefas propostas pelo tutor e visualizando os materiais disponibilizados. Já o perfil do aluno evadido pode ser entendido como o aluno que pouco interage com os recursos do Moodle e que não visualiza os materiais e as tarefas disponibilizadas.

Neste trabalho foi analisado um conjunto de dados contendo apenas dados de estudantes que concluíram o curso ou evadiram. Desta forma, os autores consideram importante a mineração de dados semestrais contendo alunos regulares e evadidos, podendo prever informações que possam levar o estudante a evadir em um determinado semestre. Nos experimentos realizados neste trabalho não foi possível utilizar esta informação por não estar disponível no conjunto de dados fornecido aos autores.

Como trabalhos futuros, também sugere-se o emprego de outras informações existentes no AVA, tais como a utilização do *quiz*, que não foi possível neste trabalho, interações pelos tutores, dados relacionados à localização dos polos de apoio e informações relacionadas a desempenho. Vale ressaltar que neste trabalho foi utilizado apenas o total de interações dos alunos nos recursos das disciplinas ofertadas no curso analisado. Por fim, com a análise qualitativa foi possível ter uma melhor interpretação dos dados e entender as características de cada experimento analisado, já que nenhum outro estudo dessa natureza foi realizado até então nesta instituição.

Referências

ABBAD, G.; CARVALHO, R. S.; ZERBINI, T. **Evasão em curso via internet: explorando variáveis explicativas**. RAE-eletrônica, v. 5, n. 2, jul./dez. 2006.

ANDERSON, T. **Toward a theory of online learning**. In: ANDERSON, T.; ELLOUMI, F. (Org.). **Theory and practice of online learning**. Athabasca: Athabasca University, 2004. p. 33-60.

BELLONI, M. L. **Educação a distância**. 2. ed. Campinas, SP: Autores Associados, 2001.

BITTENCOURT, I. M.; MERCADO, L. P. L. Evasão nos cursos na modalidade de educação a distância: estudo de caso do Curso Piloto de Administração da UFAL/UAB. **Revista Ensaio: Avaliação de Políticas Públicas em Educação**. Rio de Janeiro, v. 22, n. 83, p. 465-504, abr./jun. 2014.

BURGOS, C. et al. Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. **Computers & Electrical Engineering**. p. 1-16, mar. 2017.

CASTRO, L. N.; FERRARI, D. G. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. 1. ed. São Paulo: Saraiva, 2016.

CENSO EAD.BR: **relatório analítico da aprendizagem a distância no Brasil 2017** = Censo EAD.BR: analytic report of distance learning in Brazil 2017 [livro eletrônico]/[organização] ABED – Associação Brasileira de Educação a Distância; [traduzido por Maria Thereza Moss de Abreu]. Curitiba: InterSaberes, 2018.

CENSO EAD.BR. **Relatório analítico da aprendizagem a distância no Brasil 2018** =Censo EAD.BR: analytic report of distance learning in Brazil 2018 [livro eletrônico]/[organização] ABED – Associação Brasileira de Educação a Distância. Camila Rosa (tradutora). Curitiba: InterSaberes, 2019.

COSTA, E. *et al.* **Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações**. Jornada de Atualização em Informática na Educação, v. 1, n. 1, p. 1-29, 2013.

DIAZ, D. P.; BONTENBAL, K. F. **Learner preferences: Developing a learner-centered environment in the online or mediated classroom**. *Education at a Distance*, v. 15, n. 8, 2001.

DORE, R.; LÜSCHER, A. Z. **Permanência e evasão na educação técnica de nível médio em Minas Gerais**. *Cadernos de Pesquisa*, v. 41, n. 144, p. 772-789, 2011.

EYNG, A. M. *et al.* **Diversidade e padronização nas políticas educacionais: configurações da convivência escolar**. *Ensaio: Avaliação e Políticas Públicas em Educação*, [S.l.], v. 21, n. 81, p. 773–800, 2013.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases**. *AI Magazine*. Providence, v. 17, n. 3, p. 37-54, jul. 1996.

FAVERO, R. V. **Dialogar ou evadir: eis a questão: um estudo sobre a permanência e a evasão na educação a distância no estado do Rio Grande do Sul**. 2006. Dissertação (Mestrado) – Programa de Educação a Distância, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

GONÇALVES, O.; BELTRAME, W. (2019). **Mineração de dados e Evasão estudantil: Analisando o curso de nível superior do Ifes**. Conference: XII CASI - Congresso de Administração, Sociedade e Inovação.

MAIA, C.; J. MATTAR. **ABC da EaD: a Educação a Distância hoje**. 1. ed. São Paulo: Pearson. 2007.

MANHÃES, L. M. B. **Predição do desempenho acadêmico de graduandos utilizando mineração de dados educacionais**. 2015. 157 p. Tese (Doutorado em Engenharia de Sistemas e Computação) – Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Rio de Janeiro, 2015.

MARTÍNEZ, M.; GARCÍA M. C.; MONTORO, J. M. **Dificuldades de aprendizagem**. 1. ed. Porto: Porto editora, 2003.

MILL, D. **Mudanças de mentalidade sobre educação e tecnologia: inovações possibilidades tecnopedagógicas**. In: MILL, D. (Org.) Escritos sobre educação: desafios e possibilidades para ensinar e aprender com as tecnologias emergentes. São Paulo: Paulus, 2013.

MORAN, J. M. **O que é Educação a Distância**. Universidade de São Paulo. 2002. Disponível em: <<http://www2.eca.usp.br/moran/wp-content/uploads/2013/12/dist.pdf>>. Acesso em: 18 mar. 2019.

OLIVEIRA, J. J. G.; NORONHA, R. V.; KAESTNER, C. A. A. **Método de seleção de atributos aplicados na previsão da evasão de cursos de graduação**. Revista de Informática Aplicada. 2017.

QUEIROGA, E.; CECHINEL, C.; ARAÚJO, R. Predição de estudantes com risco de evasão em cursos técnicos a distância. In: **Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)**. Recife: Sociedade Brasileira de Computação, 2017. v. 1, p. 1547-1556.

RABELO, H. *et al.* Utilização de técnicas de mineração de dados educacionais para predição de desempenho de alunos de EaD em ambientes virtuais de aprendizagem. **Anais do SBIE 2017**, p. 1527-1536, 2017.

ROMERO, C.; VENTURA, S. **Data mining in education**. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, v. 3, n. 1, p. 12-27, 2013.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning: From theory to algorithms**. [S.l.]: Cambridge university press, 2014.

SCHMITT, J. A. **Identificação de alunos com tendência a evasão nos cursos de graduação a distância por meio de mineração de dados educacionais**. 2018. 175 p. Dissertação (Mestrado Profissional em Tecnologias Educacionais em Rede) – Universidade Federal de Santa Maria, Santa Maria, 2018.

TOCZEK, J. *et al.* **Uma visão macroscópica da evasão no ensino superior a distância do Brasil**. 2004.

ZAKI, M. J. **Parallel and Distributed Data Mining: An Introduction**. Large-Scale Parallel Data Mining. Berlin: Springer-Verlag, 2000.

COMO CITAR ESTE ARTIGO

ABNT: SONNENSTRAHL, T. S.; BERNARDI, G.; PERTILE, S. Análise de Interações do Ambiente Virtual de Aprendizagem para Predição de Evasão em Cursos no Ensino a Distância. **EaD em Foco**, v. 11, n. 1, e1463, 2021. doi: <https://doi.org/10.18264/eadf.v10i1.1463>