

Descoberta de Conhecimento em Base de Dados de Cursos de Pós-Graduação do UNA-SUS/UFCSPA para Identificação de Fatores que Contribuem na Reintegração dos Alunos

Felipe Augusto Prauchner Christmann, PPGTIG-UFCSPA,
felipe.christmann@ufcspa.edu.br, 0000-0003-3058-9146

Alessandra Dahmer, PPGTIG-UFCSPA, adahmer@ufcspa.edu.br, 0000-0003-4849-1315

Resumo: A disponibilidade de dados na educação através de cursos e plataformas online cresce cada vez mais. A oferta de cursos a distância contribui de forma significativa para a geração de dados que podem ser explorados. Este trabalho teve por objetivo analisar as características dos alunos do Curso de Especialização em Saúde da Família do UNA-SUS/UFCSPA, identificando o perfil daqueles que enfrentaram dificuldades durante curso e tiveram sucesso na sua formação. A partir das planilhas de dados dos alunos, foram selecionados os atributos de interesse e aplicado o processo de Descoberta de Conhecimento em Base de Dados com as técnicas de Árvore de Decisão e Regras de Associação. Após análise dos dados, obteve-se um conjunto de 16 perfis, 8 para os alunos reintegrados e 8 para alunos evadidos. Na análise por região, quando comparados os perfis de alunos reintegrados e evadidos, encontrou-se diferença entre todos os perfis, porém as características mais comuns foram o acesso à Internet e a presença na aula inaugural. Já na comparação dos perfis gerais (todas as regiões), estes se mostraram similares. Estes resultados podem servir para a criação de estratégias de mitigação da evasão nos cursos de ensino a distância.

Palavras-chave: Mineração de Dados. Mineração de Dados Educacionais. Educação a Distância. UNA-SUS.

Knowledge Discovery in UNA-SUS/UFCSPA Postgraduate Course Database to Identify Factors Contributing to Student Reintegration

Abstract: The availability of data in education through online courses and platforms is growing more and more. The offering of distance learning courses significantly contributes to the generation of data that can be explored. This work aimed to analyze the characteristics of students in the Specialization Course in Family Health at UNA-SUS/UFCSPA, identifying the profile of those who faced difficulties during the course and succeeded in their education. From the student data spreadsheets, the attributes of interest were selected, and the process of Knowledge Discovery in Databases was applied using Decision Tree and Association Rules techniques. After data analysis, a set of 16 profiles was obtained, 8 for reintegrated students and 8 for dropout students. In the regional analysis, when comparing the profiles of reintegrated and dropout students, differences were found among all profiles, but the most common characteristics were internet access and attendance at the inaugural class. However, in the comparison of the overall profiles (all regions), they appeared to be similar. These results can be used to create strategies to mitigate dropout rates in distance education courses.

Keywords: Data Mining, Educational Data Mining, Distance education, UNA-SUS.

1. Introdução

O uso da Internet aumenta cada dia mais em todos os setores da sociedade, seja para fazer compras em uma loja online, assistir vídeos no Youtube ou para fazer um curso na modalidade Educação a Distância (EaD) em uma instituição de ensino. Como consequência dessa popularização, a geração de dados cresceu de forma explosiva (COSTA et al. 2012). No que tange a educação, um dos responsáveis pela geração de grandes quantidades de dados é o Ambiente Virtual de Aprendizagem (AVA), presente em muitas instituições educacionais e ferramenta fundamental de ensino na atualidade, principalmente na modalidade EaD. Estas ferramentas se tornaram ainda mais relevantes durante o período pandêmico da COVID-19, onde a impossibilidade do ensino presencial obrigou as instituições de ensino a migrar para plataformas online para dar continuidade às suas atividades (DOSEA et al. 2020). A realização de um curso através de uma plataforma online exige do estudante acesso à Internet e a um computador, realidade essa que não é comum em grande parte do território nacional. Algumas regiões mais distantes dos centros urbanos sofrem com problemas de infraestrutura e acesso aos meios de comunicação e à Internet, sendo uma dificuldade extra na hora de realizar um curso na modalidade EaD. Estas questões acabam agravando os altos índices de abandono de curso que historicamente existem. Segundo Censo o EAD.BR de 2020, estima-se que até 25% dos alunos de cursos EaD evadem. Para combater tal problema, algumas soluções têm sido buscadas por gestores e/ou professores das instituições de ensino.

Uma iniciativa que se enquadra nesse cenário é a Universidade Aberta do SUS (UNA-SUS), criada em 2008 pelo Ministério da Saúde e destinada a suprir as demandas da educação contínua dos profissionais da saúde no Brasil. A partir de 2009, a Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA) se uniu a Rede UNA-SUS e inaugurou a disponibilização do seu programa de Especialização em Saúde da Família (EspSF). O acompanhamento das turmas oferecidas pela UNA-SUS/UFCSPA é feito por uma equipe pedagógica, que coleta informações relacionadas aos alunos e as utiliza para monitorar o seu andamento no curso (COSTA et al., 2015). Ao investigar estas turmas, este trabalho procura mapear o perfil do aluno através do processo de Descoberta de Conhecimento em Base de Dados (Knowledge Discovery in Databases - KDD), - que possibilita analisar conjuntos de dados para extração de conhecimento relevante - para responder a seguinte pergunta: “Como podemos caracterizar o perfil de um aluno recuperado em um curso na modalidade EaD da área da saúde, utilizando o processo de Descoberta de Conhecimento em Base de Dados?”

As técnicas de agrupamento são as mais indicadas para este caso, pois concentram as características ou objetos de estudo em grupos similares, permitindo assim que se encontre um perfil comum entre os estudantes. Outra técnica que pode ser útil são as regras de associação, que buscam identificar através da frequência de itens ou de características de um objeto se ocorre em outro objeto. Neste trabalho, os atributos que compõem o perfil servirão para fazer uma associação ao resultado do aluno ao final do curso, com foco naqueles classificados como recuperados por esta pesquisa.

Este artigo está organizado da seguinte forma: Na seção 2 são apresentados os trabalhos relacionados, na seção 3 os materiais e métodos, na seção 4 os resultados e discussões e na seção 5 as conclusões.

2. Trabalhos relacionados

A presente seção apresenta alguns trabalhos relacionados ao tema desta pesquisa, bem como traz uma breve comparação entre os métodos empregados por cada um.

As técnicas encontradas nos estudos variaram, no entanto, em sua maioria foram utilizadas Árvore de Decisão. No estudo de Manhães et al. (2011), os autores concluíram que a análise realizada é viável e que os experimentos retornaram dados com acurácia média variando entre 75% e 80%, na previsão do risco de evasão.

A pesquisa de Da Costa, Cazella e Rigo (2015), mostra que o treinamento do modelo foi realizado com um conjunto de 248 instâncias e o método de teste foi o *cross-validation*, assumindo o valor de 10 *folds*. O modelo preditivo gerado para evasão classificou corretamente 242 (97,6%) instâncias e incorretamente 6 (2,4%) instâncias.

Como resultado do estudo de Santos et al. (2015), que focava no desânimo dos alunos em relação ao curso, as Regras de Associação geradas apontaram que os alunos que se sentiram desanimados enviaram menos de 50% das atividades e tiveram dificuldades com o conteúdo, o que refletiu no conceito final.

No estudo de Paz e Cazella (2017), através dos experimentos realizados, foi verificada uma boa precisão, atingindo acurácia superior a 90%. Além disso, evidenciase a confirmação da percepção inicial de que o incentivo e o currículo dos alunos estão diretamente relacionados à tendência de evasão, e que estas ocorrem geralmente nos semestres iniciais dos cursos com alunos sem incentivo.

Em outro estudo, Do Couto e De Santana (2017) mostram como resultado que o método *Random Forest* apresentou a melhor acurácia, superior a 87%. No entanto, os autores escolheram como melhor algoritmo o classificador *Bayesian Network*, pela precisão acima de 85% e por ser capaz de possuir fácil representação de resultados, além de poder expressar o conhecimento do especialista sobre o domínio estudado.

No estudo de Rabelo et al. (2017), os autores concluíram que utilizando as técnicas e metodologia propostas, a acurácia ficou entre 93,9% e 96,5% de precisão para saber se um aluno terá ou não um desempenho satisfatório antes do final da disciplina.

Na pesquisa de Ferreira et al. (2019), a conclusão dos autores mostrou que foi possível gerar uma lista de possíveis reprovados que teriam a oportunidade de serem reintegrados antes do final da disciplina. Para a avaliação do modelo foram comparados os índices dos classificadores e destacou-se o J48 (Árvore de Decisão) com uma predominância de acurácia de 84,5%, com precisão de 85,52%.

E por fim, Rolim e Silva (2021), através da aplicação dos algoritmos KNN (*K-nearest neighbors*) e SVM (*Support Vector Machine*), obtiveram uma precisão superior a 92%, utilizando como atributos o sexo, idade em que o estudante cursou a disciplina, média final e resultado. Também foram verificados fatores determinantes associados à retenção, representados pela faixa etária dos 19 aos 23 anos, sexo e notas finais.

De modo geral, os estudos encontrados focaram na predição do risco de evasão dos estudantes com base em dados extraídos de bancos de dados do Moodle, sendo em grande parte, dados que não consideravam explicitamente o perfil dos alunos, no que diz respeito a informações socioeconômicas, idade, sexo etc. O foco da maioria dos estudos foi utilizar os dados de notas e interações com o ambiente virtual de aprendizagem, para predeterminar o desempenho acadêmico com base na mineração realizada sobre estes dados. Não foram encontrados estudos que abordavam especificamente o perfil de um aluno reintegrado, proposta do trabalho relatado neste artigo.

3. Materiais e métodos

Este trabalho é classificado como pesquisa aplicada devido a sua possível aplicação prática para as equipes de apoio dos cursos Ead. A abordagem adotada é qualitativa, uma vez que investiga uma perspectiva exploratória e subjetiva. Por fim, seu caráter é o explicativo, pois tem como foco principal retratar os dados de entrada, através

da identificação de grupos que compartilham características semelhantes que explicam o fenômeno de evasão e recuperação.

Para a realização da etapa de mineração de dados, optou-se pelo uso da ferramenta Orange Data Mining. A ferramenta Orange é escrita na linguagem de programação Python, trazendo grande versatilidade e maior performance na execução dos algoritmos. Esta ferramenta disponibiliza as suas funcionalidades através de *widgets*. Estas funcionalidades são agrupadas por categorias no canto esquerdo da tela inicial e possuem compatibilidade entre si conforme a entrada e saída de dados. Ao clicar em um *widget*, ele apresenta uma janela de configuração para que o usuário possa estabelecer os parâmetros de execução da funcionalidade, que pode ser uma simples visualização de dados até a aplicação de algoritmos.

A escolha dos algoritmos se deu através da revisão de estudos relacionados onde a Árvore de Decisão foi o algoritmo mais utilizado nas pesquisas estudadas. Devido a isso e à disponibilidade de algoritmos na ferramenta escolhida para a aplicação prática da pesquisa, constatou-se que a sua utilização se torna uma escolha viável, pois além de a Árvore de Decisão ser um algoritmo relativamente simples de aplicar, ele fornece uma fácil visualização e entendimento do problema. Para fins de complementação, definiu-se que as Regras de Associação também se caracterizam por serem algoritmos adequados ao tipo de problema desta pesquisa e foram incluídas na realização da mineração dos dados educacionais, embora não tenham sido muito utilizadas nos estudos encontrados anteriormente.

Para esta pesquisa, foram extraídos e utilizados os dados contidos em planilhas eletrônicas armazenadas em nuvem, que foram preenchidas pelo Apoio Acadêmico, Setor Pedagógico, tutores e coordenadores do projeto. Estavam disponíveis os dados de 16 turmas oferecidas a partir de 2011, organizadas em pastas com planilhas diversas, contendo informações do curso. Optou-se, por trabalhar com três turmas, totalizando uma amostra de 1125 alunos.

Após a importação dos dados para a ferramenta Orange, como etapa de pré-processamento, foram removidos todos os alunos com alguma informação faltante. Com isso, a amostra contabilizou um total de 942 alunos. Destes, 57,64% eram do sexo feminino, 51,70% eram brasileiros, 80,57% não possuíam experiência com cursos em EaD, 95% deles registraram frequência na aula inaugural, 89,92% declararam possuir acesso à Internet em casa ou no trabalho e 70,81% foram aprovados ao final do curso. Os alunos eram residentes dos seguintes estados do Brasil: Acre (3,5%), Amazonas (17,94%), Amapá (2,12%), Pará (29,29%), Roraima (3,08%), Rio Grande do Sul (38,32%) e Sergipe (5,84%).

Dois atributos novos foram criados a partir dos dados existentes. O primeiro chamado RESULTADO, fruto da combinação do atributo CONTATO APOIO e RECUPERAÇÃO para formar o principal perfil desta pesquisa: Aluno Reintegrado (AR). O aluno que foi considerado reintegrado possui os valores “S” em RECUPERAÇÃO e “S” em CONTATO APOIO. Sendo assim, o atributo original DESFECHO foi reclassificado e copiado para um novo atributo chamado RESULTADO, ficando da seguinte forma: Aprovado Normalmente (AP), Aprovado Reintegrado (AR), Reprovado (RP) e Evadido (E). O aluno que seguiu o fluxo normal do curso foi classificado com AP, o aluno que precisou recuperar alguma nota e teve algum contato por parte do Apoio foi classificado com AR, o aluno reprovado por nota foi classificado com RP e, por fim, o aluno que foi desligado do curso ou o abandonou foi classificado com E. Deve-se destacar que um dos objetivos deste trabalho foi o de definir como obter o perfil do aluno reintegrado, resultando na criação desse atributo. O segundo atributo criado chama-se INTERNET resultado de uma simplificação dos dois atributos relacionados ao acesso à

Internet. Foi considerado que o aluno que tivesse internet em casa ou no trabalho possuía acesso à Internet.

Neste trabalho estamos interessados em saber se alguma das características do perfil do aluno aparece com frequência nos casos em que o aluno foi marcado como AR e compará-lo com os perfis dos alunos evadidos.

Utilizando o algoritmo Regras de associação, obteve-se o seguinte resultado:

Quadro 1 – Regras resultantes para os alunos reintegrados.

Suporte	Confiança	Antecedente	Consequente
0,006	0,6	UF=AM, EXPERIÊNCIA EM EaD?=N, SEXO=M, AULA INAUGURAL=S, AULA EIXO 2=N	RESULTADO=AR
0,005	0,625	NACIONALIDADE=B, UF=RR, EXPERIÊNCIA EM EaD?=N, SEXO=F	RESULTADO=AR
0,005	0,625	NACIONALIDADE=B, UF=RR, EXPERIÊNCIA EM EaD?=N, SEXO=F, INTERNET=S	RESULTADO=AR
0,005	0,625	NACIONALIDADE=B, UF=RR, EXPERIÊNCIA EM EaD?=N, SEXO=F, AULA INAUGURAL=S	RESULTADO=AR
0,005	0,714	UF=RR, EXPERIÊNCIA EM EaD?=N, INTERNET=N, AULA INAUGURAL=S	RESULTADO=AR
0,005	0,625	NACIONALIDADE=B, UF=RR, EXPERIÊNCIA EM EaD?=N, SEXO=F, INTERNET=S, AULA INAUGURAL=S	RESULTADO=AR

Os parâmetros utilizados na ferramenta Orange foram: suporte mínimo de 5% e confiança mínima de 60%. Além disso, as regras foram filtradas por consequente na forma Y=AR e que contenha no mínimo 4 antecedentes. Como resultado, obtém-se 6 regras de associação. Neste caso, podemos destacar dois perfis de alunos com maior confiança para cada estado que apareceram no resultado: com confiança de 71,4% e suporte de 5%, o perfil do aluno do estado de Roraima, sem experiência em EaD, sem acesso à Internet e que participou da aula inaugural. O outro perfil que podemos destacar possui confiança de 60% e suporte de 5% e pertence ao aluno do estado do Amazonas, sem experiência em EaD, do sexo masculino e que participou da aula inaugural, mas não participou da aula do Eixo 2. Estes serão os perfis comparados com o aluno evadido na parte de resultados deste trabalho.

Aplicando as Regras de Associação e escolhendo como alvo os alunos evadidos, temos o seguinte resultado:

Quadro 2 – Regras resultantes para os alunos Evadidos.

Suporte	Confiança	Antecedente	Consequente
0,007	0,333	UF=RS, NACIONALIDADE=B, AULA EIXO 2=N	RESULTADO=E
0,007	0,333	UF=RS, NACIONALIDADE=B, AULA EIXO 2=N, INTERNET=S	RESULTADO=E
0,006	0,500	AULA INAUGURAL=N, AULA EIXO 2=N, INTERNET=S	RESULTADO=E
0,005	0,333	NACIONALIDADE=B, EXPERIÊNCIA EM EaD?=S, AULA EIXO 2=N, INTERNET=S	RESULTADO=E
0,005	0,312	UF=RS, NACIONALIDADE=B, AULA EIXO 2=N, AULA INAUGURAL=S	RESULTADO=E
0,005	0,312	NACIONALIDADE=B, UF=RS, INTERNET=S, AULA INAUGURAL=S, AULA EIXO 2=N	RESULTADO=E

Os parâmetros utilizados foram: suporte mínimo de 5% e confiança mínima de 30%. Além disso, as regras foram filtradas por consequente ($Y=E$) e que contenham no mínimo 3 antecedentes. Como resultado, obtém-se 6 regras de associação.

Sendo assim, podemos destacar o perfil do aluno do estado do Rio Grande do Sul, sem acesso à internet e que não participou de nenhuma aula presencial. Esta regra de associação possui confiança de 50% e suporte de 5%.

O outro algoritmo aplicado foi Árvores de decisão, que se caracteriza por serem estruturas visuais que têm formato de uma árvore, onde a raiz contém os objetos a serem classificados e as ramificações (nós) são atributos de semelhança destes objetos. Os objetos vão percorrendo os nós até chegarem na folha (final do caminho), que vai definir a classificação deste objeto.

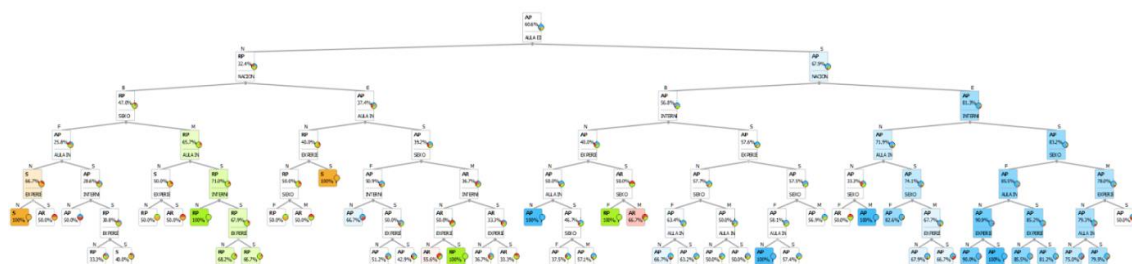


Figura 1 – Árvore de Decisão para toda a amostra.

A amostra foi então separada por estado para ser apresentada e analisada na próxima seção. Os perfis serão apresentados nos quadros 4 e 5.

4. Resultados e discussão

Através do uso das técnicas de Mineração de Dados, buscou-se entender o perfil do aluno em dificuldades no Curso de Especialização em Saúde da Família oferecido pela UNA-SUS/UFCSPA, que conseguiu terminar o curso com sucesso e compará-lo com os alunos que foram desligados do curso por algum motivo. Esta comparação visa entender se houve fatores que influenciaram na recuperação deste aluno, bem como se podemos atribuir a atuação do Apoio Acadêmico estes casos de sucesso.

É importante destacar que alguns atributos desta amostra estavam bastante desbalanceados, ou seja, um único valor aparecia na maioria das instâncias, como é o caso dos atributos AULA INAUGURAL, AULA EIXO 2, EXPERIÊNCIA EaD e INTERNET. Apesar de existirem técnicas para lidar com este tipo de situação, optou-se por não se aplicar nenhuma técnica de balanceamento para não haver interferência nas amostras e não diminuir ainda mais o número de instâncias.

A aplicação do algoritmo de Regras de Associação resultou em duas regras principais para o aluno reintegrado e uma regra para o aluno evadido, apresentadas no quadro abaixo:

Quadro 3 – Regras resultantes após seleção das Regras de Associação.

Suporte	Confiança	Antecedente	Consequente
0,006	0,6	UF=AM, EXPERIÊNCIA EM EaD?=N, SEXO=M, AULA INAUGURAL=S, AULA EIXO 2=N	RESULTADO=AR
0,005	0,714	UF=RR, EXPERIÊNCIA EM EaD?=N, INTERNET=N, AULA INAUGURAL=S	RESULTADO=AR
0,005	0,333	INTERNET=S, AULA INAUGURAL=N, AULA EIXO 2=N	RESULTADO=E

A regra com a maior confiança diz que o aluno de Roraima sem experiência em cursos EaD, sem acesso à Internet e que participou da aula inaugural foi reintegrado pelo Apoio em 71,4% dos casos.

A regra com a segunda maior confiança entre as geradas mostra que o aluno do Amazonas teve características semelhantes aos alunos de Roraima, resultando em 60% dos casos reintegrados pelo Apoio Acadêmico, segundo o critério desta pesquisa.

Os dois estados brasileiros que apareceram nas Regras de Associação mineradas nesta pesquisa se caracterizam por serem estados precários em infraestrutura e de difícil acesso as localidades (COSTA et al., 2015). Vale destacar que em casos em que o aluno tinha dificuldades em acessar o material do curso pela Internet, a equipe do UNA-SUS/UFCSPA enviava o material de estudos gravados em CD's e *pendrives* para que o aluno pudesse estudar e quando tivesse oportunidade de acessar à Internet, poderia fazer os exercícios de forma mais tranquila (COSTA et al., 2015). Este fato pode explicar o aparecimento destas características neste caso.

Os resultados obtidos com o algoritmo de Árvores de Decisão são apresentados através dos seguintes quadros.

Quadro 4 – Comparação dos perfis de alunos reintegrados.

ATRIBUTO	RESULTADO: REINTEGRADO							
	AC	AM	AP	PA	RR	RS	SE	BRASIL
Sexo	M	M	F	M	F	F	M	M
Nacionalidade	E	B	E	E	B	E	B	B
Experiência em EAD	N	N	N	N	N	N	N	N
Aula inaugural	S	S	N	S	S	S	S	S
Aula Eixo 2	S	S	S	S	S	N	S	S
Acesso à Internet	S	S	S	S	S	S	S	S

Quadro 5 – Comparação dos perfis de alunos evadidos.

ATRIBUTO	RESULTADO: EVADIDO							
	AC	AM	AP	PA	RR	RS	SE	BRASIL
Sexo	F	F	*	M	*	M	F	M
Nacionalidade	E	B	*	B	*	B	B	B
Experiência em EAD	N	N	*	N	*	S	S	N
Aula inaugural	S	N	*	S	*	S	S	S
Aula Eixo 2	N	N	*	S	*	S	N	S
Acesso à Internet	S	S	*	S	*	S	S	S

No estado do Acre, três atributos se diferenciam entre os perfis: SEXO, NACIONALIDADE e AULA EIXO 2. No caso do atributo SEXO, faltam informações para se chegar a alguma conclusão baseada nos dados investigados nesta pesquisa. Em relação ao outro atributo, a ausência do aluno evadido na segunda aula presencial pode significar que este aluno se desligou cedo do curso, não permitindo uma maior atuação do Apoio Acadêmico. Como no perfil do aluno reintegrado ele estava presente na segunda aula presencial, isso pode ter influenciado na sua recuperação. Este estado também se caracteriza por ser de difícil acesso e possuir uma infraestrutura precária (COSTA et al., 2015), onde apesar de os alunos terem acesso à internet, a pesquisa mostrou que esse pode não ser um fator relevante na recuperação do aluno. No estado do Amazonas, chama a atenção a ausência dos alunos evadidos nas duas aulas presenciais, mostrando que este aluno nem ao menos começou o curso. Vários fatores podem ter contribuído para esta

realidade, como por exemplo, o difícil acesso ao local da aula presencial. Para o aluno reintegrado, apesar dos dados sugerirem que ele esteve presente em todo o curso, outros fatores podem ter influência na dificuldade deste aluno e que a atuação do apoio acadêmico pode ter suprido essa necessidade.

No estado do Amapá, não houve casos de alunos evadidos, portanto não foi possível fazer um comparativo entre os perfis dos alunos reintegrados e evadidos. Em relação ao perfil dos alunos reintegrados, vale destacar que a ausência na aula inaugural e presença na aula do Eixo 2 mostra que já pode ter havido uma atuação do Apoio Acadêmico para a reintegração desse aluno no curso. Infelizmente este estado teve poucas instâncias, prejudicando a extração de conhecimento com as técnicas utilizadas.

No estado do Pará, houve diferença apenas na nacionalidade do aluno, onde o perfil do aluno reintegrado é estrangeiro. Neste caso, os dados não fornecem informações suficientes para uma conclusão. Uma possível explicação é a questão de adaptação dos estrangeiros a realidade brasileira.

No estado de Roraima, novamente não houve casos de alunos evadidos, sendo assim não foi possível fazer um comparativo entre os perfis dos alunos reintegrados e evadidos. O estado de Roraima possuía uma amostra muito pequena de alunos reintegrados, tornando essa análise em particular também inconclusiva. No entanto, este foi um dos estados que apareceu na mineração de dados utilizando a técnica de Regras de Associação, o que pode ser explicado pelo pequeno número de instâncias, já que relativamente, cada indivíduo tem uma influência maior sobre as outras instâncias.

O estado do Rio Grande do Sul mostra um perfil atípico para o aluno evadido, já que as características sugerem que este aluno teve condições de concluir o curso. Este perfil pode indicar que o aluno desistiu perto do final do curso. Já o perfil do aluno reintegrado pode indicar que o Apoio atuou para manter o aluno no curso, apesar da ausência na segunda aula presencial.

Para o estado de Sergipe, percebe-se que o perfil do aluno reintegrado participou das duas aulas presenciais, o que mostra que a dificuldade talvez estivesse em outros fatores que não somente as notas baixas ou atrasos de entrega. Já o perfil do aluno evadido que apesar de possuir experiência em cursos a distância, os estudantes já haviam desistido do curso antes do segundo módulo.

Quando analisamos os perfis sem a distinção por estado, denominado Brasil, o resultado do perfil emergido foi o mesmo para os alunos reintegrados e evadidos. Este perfil é igual ao emergido nos estados de Sergipe e Amazonas, no caso dos alunos reintegrados e no caso dos alunos evadidos, este mesmo perfil apareceu no estado do Pará. Pode-se concluir que houve uma grande variedade nos perfis emergidos, mostrando que os fatores regionais podem ser também determinantes ou que os dados não foram consistentes o suficiente para convergir o perfil do aluno. No caso dos alunos evadidos, dois estados não tinham nenhuma instância, pois o banco de dados contava com apenas 44 ocorrências deste tipo de aluno, dificultando a comparação dos resultados. Isso pode ter se dado pela obrigatoriedade da inscrição no curso, que vinculava a conclusão à continuidade nos programas Mais Médicos e PROVAB. Outrossim, observa-se que o atributo de acesso à Internet foi o único presente em todos os perfis, provavelmente pelo desbalanceamento deste atributo dentro da amostra de dados.

Com estes resultados não foi possível obter informações sólidas a respeito das características dos alunos que participaram do curso EspSF. No entanto, podem ser úteis para direcionar trabalhos futuros e melhorias na coleta e tratamento dos dados, conforme limitações apontadas durante a pesquisa.

5. Conclusões

Em resumo, a grande maioria dos alunos possuía acesso à Internet, o que mostrou que este fato sozinho não explica a diferença entre um aluno aprovado e um evadido, como era de se esperar. A falta de experiência anterior com cursos a distância parece também não ser um fator decisivo para essa diferenciação. A frequência nas aulas presenciais, o gênero e o fato de o aluno ser estrangeiro ou brasileiro foram os fatores que mais influenciaram na diferenciação dos perfis.

A hipótese norteadora da pesquisa era de que a metodologia utilizada no curso Curso de EspSF do UNA-SUS/UFCSPA através de uma equipe de Apoio Acadêmico (COSTA et al., 2015) seria muito importante para a realização de cursos na modalidade EaD, ainda mais em se tratando de públicos tão abrangentes como foi o caso. A mineração dos perfis pode permitir uma aferição das ações que estão funcionando e permite a criação de novas iniciativas para auxiliar os alunos a terminarem o curso. O perfil de aluno reintegrado é de suma importância neste contexto, pois mostra as características que devem ser observadas na hora de se pensar as estratégias para evitar a alta evasão. Apesar de não haver uma conclusão a respeito de um perfil único do aluno, a definição de aluno reintegrado pode ser utilizada em trabalhos futuros para a obtenção deste perfil.

A comparação dos perfis reintegrados com os perfis evadidos pode mostrar que, em caso de diferença entre eles, as características evidenciadas podem elucidar sobre a abordagem das ações do Apoio Acadêmico. No caso de emergirem os mesmos perfis, a explicação da diferença entre evadidos e reintegrados não se encontra nos atributos escolhidos ou os dados não foram suficientes para evidenciar essa diferenciação.

Em relação as limitações da pesquisa, o escopo dos dados ficou limitado a três turmas do curso de especialização. Além disso, não houve uma diferenciação entre a origem dos comentários do atributo ANOTAÇÕES APOIO, que tiveram origem em diferentes setores do curso e não apenas do Apoio Acadêmico e que foi fundamental na definição do perfil do aluno reintegrado. Também não foi definida nenhuma ação de apoio ao perfil de aluno descoberto, podendo ser explorado em trabalhos futuros.

No que tange a uma possível continuidade deste trabalho, a pesquisa pode ser estendida com uma maior quantidade de turmas do UNA-SUS/UFCSPA e com a escolha de outros atributos disponíveis nas planilhas e outras fontes de dados. A nota dos alunos também é um material que não foi muito explorado nesta pesquisa e pode fornecer novas informações sobre o perfil do aluno que frequentou este curso.

Referências

- ABED - ASSOCIAÇÃO BRASILEIRA DE EDUCAÇÃO A DISTÂNCIA. **Censo EAD.BR 2020 - Relatório analítico da aprendizagem a distância no Brasil.** Disponível em: < https://abed.org.br/arquivos/CENSO_EAD_2020_PORTUGUES.pdf >. Acesso em: 2 jul. 2023.
- AGGARWAL, C. C. **Data Mining: The Textbook.** Springer International Publishing, 2015.
- COSTA, E. et al. Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. **Jornada de Atualização em Informática na Educação - JAIE 2012**, v. d, p. 1–29, 2012.
- COSTA, M. R. et al. **Encurtando distâncias: o papel do apoio acadêmico em um curso de especialização em saúde da família.** Disponível em: < https://ares.unasus.gov.br/acervo/html/ARES/24803/1/Encurtando%20a%20dist%C3%A2ncia_Cap2.pdf >. Acesso em: 2 jul. 2023.

- DA COSTA, S. S.; CAZELLA, S.; RIGO, S. J. Minerando dados sobre o desempenho de alunos de cursos de educação permanente em modalidade EAD: Um estudo de caso sobre evasão escolar na UNA-SUS. **Renote**, v. 12, n. 2, 2015.
- DO COUTO, D. DA C.; DE SANTANA, A. L. **Análise dos problemas da Evasão e Retenção - Uma abordagem através de Mineração de Dados Educacionais**. XIII Encontro Anual de Computação - EnAComp 2017. Universidade Federal de Goiás, 2017.
- DOSEA, G. S. et al. MÉTODOS ATIVOS DE APRENDIZAGEM NO ENSINO ONLINE: A OPINIÃO DE UNIVERSITÁRIOS DURANTE A PANDEMIA DE COVID-19. **Interfaces Científicas - Educação**, v. 10, n. 1, p. 137–148, 6 set. 2020.
- FERREIRA, J. L. C. et al. **A model for predicting disapproval of apprentices in distance education using decision tree**. ACM International Conference Proceeding Series. Association for Computing Machinery, 20 maio 2019.
- MANHÃES, M. L. B. et al. **Previsão de Estudantes com Risco de Evasão Utilizando Técnicas de Mineração de Dados**. Anais do XXII SBIE - XVII WIE. 21 a 25 nov. 2011.
- BRASIL; MINISTÉRIO DA SAÚDE. **UNA-SUS - Universidade Aberta do SUS**. Disponível em: <<https://www.unasus.gov.br/>>. Acesso em: 2 jul. 2023.
- PAZ, F.; CAZELLA, S. **Identificando o perfil de evasão de alunos de graduação através da Mineração de dados Educacionais: um estudo de caso de uma Universidade Comunitária**. Anais dos Workshops do VI Congresso Brasileiro de Informática na Educação (CBIE 2017). Brazilian Computer Society (Sociedade Brasileira de Computação - SBC), 10 nov. 2017.
- 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 XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017), v. 1, n. Cbie, p. 1527, 2017a.
- ROLIM, J. P.; SILVA, R. C. **Mineração de Dados Educacionais para identificação de Perfil de Retenção em um curso de Ciência da Computação** Em: ESCOLA REGIONAL DE INFORMÁTICA DE GOIÁS (ERI-GO), 9. , 2021, Evento Online. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2021 . p. 195-204.
- SANTOS, F. D. et al. **Análise de Evidências do Estado de Ânimo Desanimado de Alunos de um AVEA: uma Proposta a partir da Aplicação de Regras de Associação**. Anais dos Workshops do IV Congresso Brasileiro de Informática na Educação (CBIE 2015). Sociedade Brasileira de Computação - SBC, 26 out. 2015.
- SOCIEDADE INTERNACIONAL DE MINERAÇÃO DE DADOS EDUCACIONAIS. **Educational Data Mining**. Disponível em: <<https://educationaldatamining.org/>>. Acesso em: 2 jul. 2023.