

APEU - Um Assistente de Predição de Evasão Universitária para o Instituto Federal Farroupilha *campus* Panambi

Victoria B. Martins¹, Ederson Bastiani², Renan L. de Souza³

^{1, 2, 3} Instituto Federal Farroupilha (IFFar) *campus* Panambi - RS

victoria.martins@aluno.iffar.edu.br, {ederson.bastiani,
renan.souza}@iffarroupilha.edu.br

Abstract. *This article presents the development of a evasion system prediction for students in the Instituto Federal Farroupilha campus Panambi using data from students that already dropped out or completed their degree. This study uses the concept of Educational Data Mining (EDM). To comprehend how brazilian educational institutions are predicting probabilities of evasion using Artificial Intelligence, a System Mapping was made. Using this methodology, it was possible to define the software (WEKA) and the AI algorithm (ANN Multilayer Perceptron). After the students' data gathering, the data was processed using the Knowledge Data Discovery technique. The system was capable of 66,6% of accuracy.*

Resumo. *Este artigo apresenta o desenvolvimento de um sistema de predição de evasão de alunos do Instituto Federal Farroupilha campus Panambi baseado em perfis de alunos que já evadiram ou concluíram o curso utilizando o conceito de Mineração de Dados Educacionais (MDE). Para tanto, foi realizado um mapeamento sistemático a fim de compreender como as instituições de ensino brasileiras estão predizendo a probabilidade de evasão de seus alunos através de Inteligência Artificial. Com este estudo, foi possível definir o software (WEKA) e o algoritmo de IA (RNA Multilayer Perceptron) a ser utilizado. Após coletados os dados de alunos da instituição, o processamento destes dados foi realizado utilizando a técnica de Knowledge Data Discovery. A ferramenta chegou a aproximadamente 66,6% de acurácia.*

1. Introdução

A democratização das universidades públicas têm cada vez mais atraído indivíduos para as instituições de ensino superior. Os Institutos Federais de Educação, Ciência e Tecnologia (IFs), por exemplo, possuem como objetivo ofertar educação profissional e tecnológica, em todos os seus níveis e modalidades, formando e qualificando cidadãos com vistas na atuação profissional nos diversos setores da economia [Brasil 2008].

Democratizar, entretanto, não significa apenas garantir educação para todos, mas garantir a permanência do aluno durante seu período de estudo, oportunizando-lhe ensino de qualidade e subsídios para sua estadia [Coelho 2013]. Nesse sentido, torna-se

preocupante a crescente taxa de evasão escolar, um fenômeno social complexo definido como interrupção no ciclo de estudos [Baggi e Lopes 2011 apud Gaiosio 2005].

A evasão se configura como um obstáculo resultante de uma gama de fatores sociais, culturais e financeiros que, além de acumular prejuízos econômicos, também evocam problemas para a instituição e a sociedade [Rodriguez 2011] que perde um profissional-aluno com potencial e o próprio discente evadido, que abstrai-se da possibilidade de futuro diferente. Por estas consequências, tem sido um tema recorrente na academia pois em setores públicos são recursos públicos investidos sem o devido retorno e no setor privado são perdas de receitas [Filho et al. 2007].

Segundo a Plataforma Nilo Peçanha (2018), os IFs, incluindo os Centros Federais de Educação Tecnológica (CEFETs), possuíram taxa de evasão de 14.5% em 2018. Delimitando ao contexto deste estudo, a mesma plataforma informa que o Instituto Federal Farroupilha (IFFar) *campus* Panambi, apresentou uma taxa de evasão para os cursos de graduação de 32,9% em 2018, um crescimento de 4,7% em relação ao ano anterior.

Apesar de nem toda evasão poder ser evitada, esta deve ser estudada e compreendida para que seja possível distanciar as evasões e direcionar alunos com base em suas expectativas, necessidades e dificuldades. Desta forma, as instituições poderiam - e deveriam - conciliar sua administração com tecnologias capazes de auxiliar a combater o fenômeno da evasão, aliando conhecimentos técnicos e tecnológicos às suas práticas pedagógicas para com a sociedade brasileira.

Uma das formas de compreender esta problemática é utilizando conceitos de Mineração de Dados Educacionais (MDE). A MDE pode ser definida como a aplicação de Mineração de Dados (descoberta de conhecimento) na área da educação a fim de descobrir informações que possam auxiliar no contexto educacional [Candão e Real, 2018], tais como previsão de desempenho de alunos [Nascimento, Junior e Fagundes 2018].

Nesse sentido, pesquisadores têm reunido esforços em torno desta problemática. O trabalho de Paz e Cazella (2017), por exemplo, buscou identificar o perfil de evasão de alunos de graduação utilizando mineração de dados em uma universidade comunitária. Já Calixto, Segundo e Gusmão (2017), tiveram como foco a identificação das variáveis, tanto de alunos quanto de instituição, que tornam a evasão propensa em dois diferentes estados brasileiros. Araújo (2018), por sua vez, focou em um sistema de Mineração de Dados voltado à tomada de decisões. Ainda, Medeiros e Padilha (2018), fizeram um estudo acerca da utilização de algoritmos de classificação para detecção de a evasão escolar. Por fim, Coutinho et. al (2018) conseguiram aliar Visualização e Mineração de Dados para analisar a evasão em cursos de graduação.

Observando as potencialidades das técnicas de MDE, o problema apresentado e o movimento acadêmico em torno destes temas, este estudo apresenta uma proposta de ferramenta *web*, voltada para o IFFar *campus* Panambi, que possibilita a predição de probabilidade de evasão de alunos com base em perfis de alunos já formados e evadidos através de seus dados socioeconômicos. Sua aplicação em setores de gestão acadêmica pode auxiliar gestores no desenvolvimento de instrumentos pedagógicos de prevenção a evasão, como projetos ou oficinas. Ainda, esta proposta, pode permitir a extração de

padrões em diferentes turmas e semestres do *campus*, buscando definir, por exemplo, se o perfil de evasão é o mesmo para turmas distintas.

Desta forma, espera-se colaborar com uma solução tecnológica e científica para tomada de decisão a respeito da problemática da evasão escolar, visto que a instituição não possui sistemas com base de dados para controle e análise dos casos de desistência dos alunos no *campus*, o que dificulta a compreensão, administração, estudo e possível diagnóstico para a questão apresentada.

Este artigo está estruturado da seguinte maneira: a seção 2 apresenta as metodologias utilizadas para a realização deste trabalho. A seção 3 apresenta os resultados e, por sua vez, a seção 4 as conclusões. E por fim, apresentam-se os agradecimentos e as referências bibliográficas.

2. Metodologia

Esta seção apresenta a metodologia utilizada no desenvolvimento deste trabalho, a qual foi estruturada em três etapas. A primeira etapa deste trabalho consistiu em realizar um mapeamento sistemático de literatura, de modo a compreender um pouco mais sobre o estado da arte das técnicas de refreamento de evasão nas instituições de ensino brasileiras, utilizando a metodologia de Mineração de Dados Educacionais. Após, realizou-se o processo de MDE, a fim de compreender mais a fundo quais os fatores podem causar desistência a partir dos dados angariados. Por fim, foi desenvolvida uma ferramenta que calcula a probabilidade de evasão a partir destes dados. Estas etapas são descritas a seguir.

2.1 Mapeamento Sistemático de Literatura

Um mapeamento sistemático (MS) é uma forma de identificar, avaliar e interpretar todas as pesquisas relevantes disponíveis para uma questão de pesquisa particular [Santos e Barreto, 2015]. Nesse sentido, foi realizado um MS com o objetivo de compreender quais são formas de redução de evasão aliadas à MDE que estão sendo estudadas pelas instituições de ensino brasileiras.

Para esta pesquisa foram definidas as seguintes etapas, baseadas na metodologia proposta por Kitchenham (2004): planejamento da pesquisa; seleção de estudos primários; avaliação da qualidade; extração de dados; e análise dos resultados. A Figura 1 representa o processo do MS. A seguir apresenta-se o percurso da concepção e execução da pesquisa tendo em vista a metodologia adotada.

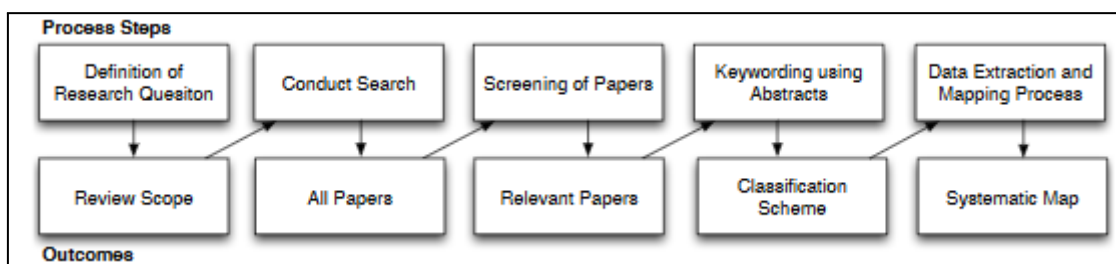


Figura 1 - Processo do Mapeamento Sistemático. Fonte: Petersen et al (2008)

Como parte da metodologia, foi definida a questão de pesquisa “Como um *software* baseado em Inteligência Artificial pode prever a possibilidade de evasão de

alunos ingressantes do IFFAR a partir de seus perfis?”, de forma a compreender: (1) quais tipos de dados têm sido obtidos; (2) quais recursos computacionais têm sido utilizados; e (3) quais algoritmos têm sido aplicados.

Para o processo de busca dos trabalhos, de forma a compreender o contexto nacional, foram utilizadas como fontes o *Portal de Publicações da Comissão Especial de Informática da Educação*¹, o qual inclui anais dos principais eventos do país nesta área, e *Revista Novas Tecnologias na Educação*². Para cada fonte, a seguinte *string* de busca foi aplicada:

“(learning analytics OR mineração de dados OR mineração de dados educacionais) AND (evasão OR desistência escolar OR abandono)”

O motor de busca em ambos as fontes de busca são idênticos, compreendendo termos como AND, OR e NOT e não diferenciando maiúsculas de minúsculas, portanto não houve diferenciação no processo de procura. A partir das combinações de palavras, por exemplo “learning analytics AND evasão”, depois “learning analytics AND desistência escolar” e assim sucessivamente, foram feitas buscas manuais. Ainda, para cada fonte, optou-se por buscar os trabalhos que apresentasse os termos da *string* de busca no título ou no resumo. Nesta etapa, foram obtidos 42 artigos publicados nos últimos 5 anos (critério de inclusão).

Após, foram excluídos os trabalhos duplicados, revisões ou mapeamentos secundários ou escritos em idiomas diferentes de Português ou Inglês. Em seguida, foram aplicados outros dois critérios de exclusão, removendo-se os trabalhos onde os termos da *string* de busca não aparecem no resumo e/ou no título e não apresentam metodologia e resultados. Ao final, 11 (26,1%) trabalhos foram identificados como relevantes para este estudo. Sucessivamente à seleção de estudos, procedeu-se a extração de dados essenciais para as respostas das questões de pesquisa³.

Quanto às ferramentas computacionais, percebeu-se que a maioria dos trabalhos utilizou WEKA, seguindo de SPSS, RStudio e RProject. Ainda 4 trabalhos não citaram os recursos utilizados para compreender os dados, apenas havendo citações sobre a base de dados e seu desenvolvimento. Pôde-se notar que artigos que buscavam analisar dados socioeconômicos abundantes utilizaram SPSS. A Figura 2 apresenta o quantitativo de trabalhos em razão das ferramentas computacionais utilizadas.

¹ <https://www.br-ie.org/pub/index.php/index/index>

² <https://seer.ufrgs.br/renote/index>

³ A tabulação dos dados para este mapeamento sistemático estão disponíveis em: <https://docs.google.com/spreadsheets/d/18iHRf0Nf8P6o7E8HgOh8UeREmYF-KldFTrznV9dbvls>

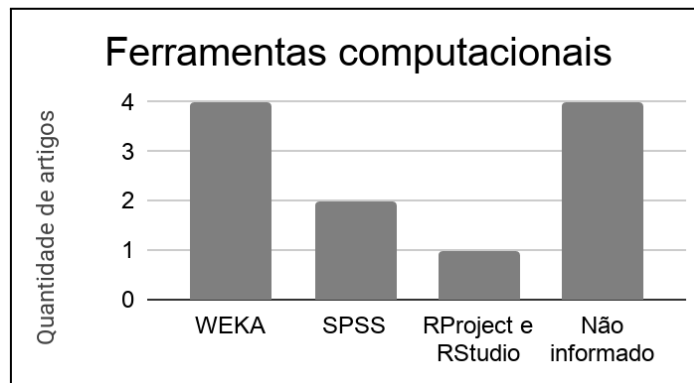


Figura 2 - Gráfico de distribuição de ferramentas computacionais

Com relação a algoritmos, notou-se que os artigos se concentraram em Árvores de Decisão e Redes Neurais Artificiais (RNA). Além de Regressões Logística (R. Log) ou Linear (R. Lin) são abordadas em 4 trabalhos, como pode ser visto na Figura 3.

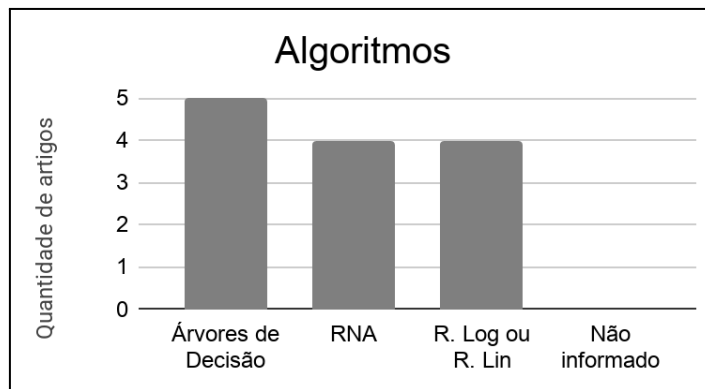


Figura 3 - Gráfico de distribuição de algoritmos

Sobre a Figura 3, é importante salientar que alguns trabalhos abordam mais que um algoritmo.

Por fim, quanto aos tipos de dados considerados pelos artigos, pode-se categorizá-los como fontes virtuais, dados socioeconômicos ou escolares. Aqueles referidos como fontes virtuais, utilizaram interações on-line como número de acessos, interações (fóruns e chats) e notas ou situação final do aluno. Enquanto os de dados socioeconômicos, colheram dados do aluno, censos escolares, envolvendo dados a respeito da estrutura escolar ou ainda, dados mais abrangentes como a base de dados do INEP. A Figura 4 apresenta o resultado quantitativo para esta questão.

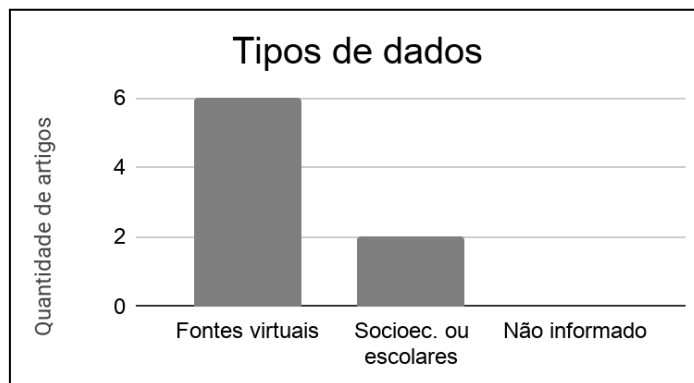


Figura 4 - Gráfico de distribuição de tipos de dados

2.2. Etapa de Mineração de Dados Educacionais

A metodologia utilizada para o processo de MDE foi a de *Knowledge Discovery from Data* (KDD), que, segundo Han, Kamber e Pei (2012), é a geração de conhecimento a partir da extração de padrões de repositórios de dados. Ainda, de acordo com os autores, consiste nas seguintes etapas: limpeza de dados, integração de dados, seleção de dados, transformação de dados, mineração de dados, avaliação de padrões e apresentação de conhecimento, como mostra a Figura 5.

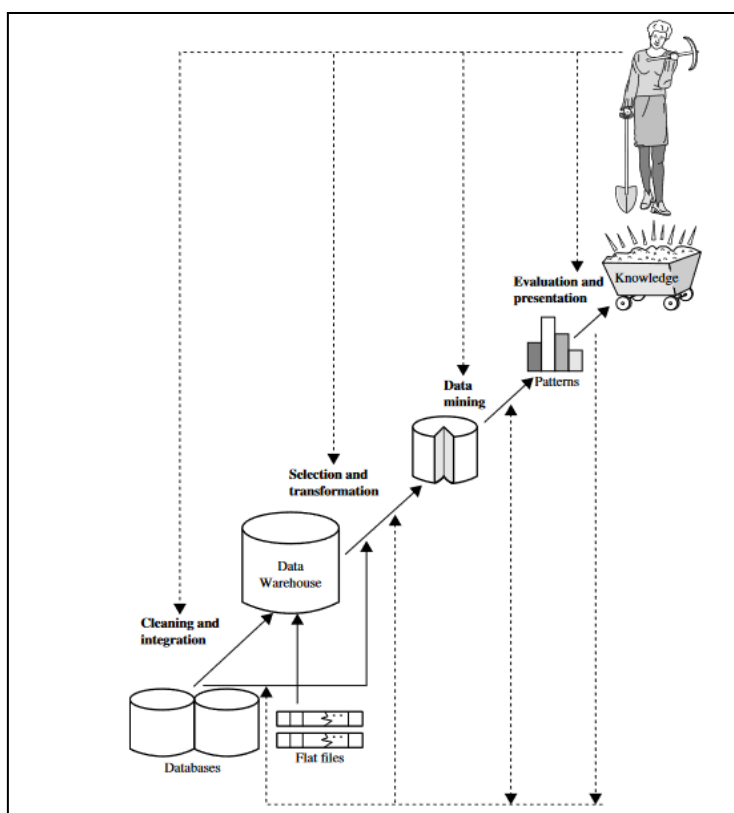


Figura 5 - Representação do Processo de KDD. Fonte: Han, Kamber e Pei, 2012.

A obtenção de dados foi realizada através do método de fichamento por meio de documentos de requerimento de matrícula de alunos do Curso Superior de Tecnologia

em Sistemas para Internet dos últimos 6 semestres (2017/1º semestre a 2019/1º semestre). Estes documentos foram disponibilizados pela Coordenação de Registros Acadêmicos do *campus*, sem a revelação das identidades dos alunos. Foram, também, considerados os alunos que concluíram o curso com até um semestre a mais do que o esperado. A Figura 6 apresenta parte dos dados fornecidos pelo setor. Valores representados por asteriscos (*) representam dados não preenchidos.

Status	Data Nasc	Estado Civil	Raça	Município	Forma Ingresso	Renda per c	Trabalho	Reserva de
Concluído este ano	22/12/1983	Solteiro	Branca	Panambi	Enem/Sisu	b	b	AC
Concluído este ano	22/7/1998	Solteiro	Branca	Panambi	Enem/Sisu	b	a	ep<=1.5
Concluído	25/6/1996	Solteiro	Branca	Panambi	Enem/Sisu	b	a	AC
Concluído	22/9/1985	Solteiro	Branca	Panambi	Enem/Sisu	*	b	ep>1.5

Figura 6. Representação parcial dos dados angariados através dos documentos disponibilizados. Fonte: autores.

Através da técnica de descarte apresentada por Assunção (2012), os documentos com valores ausentes não foram considerados, totalizando 27 registros de alunos.

Após a tabulação dos dados iniciais, realizou-se a transformação de dados numéricos sequenciais em intervalos fechados, como o intervalo de idade e de distância. Devido a grande variação de idades [SPSS, sem data], o intervalo de idade foi definido como atributo através da distribuição de frequência destas, que foram calculadas pela diferença entre a data de nascimento e o primeiro dia do mês da matrícula (março) do ano de ingresso. O processo semelhante foi realizado para o intervalo de distância, sendo que esta foi estabelecida pela diferença entre a distância do município de Panambi com a cidade referida. Caso o município seja Panambi a distância é classificada como zero. Ainda, os valores do atributo trabalho foram reduzidos de 6 opções do documento oficial para 2 valores (trabalha ou não trabalha).

Assim, definiu-se relevante a identificação dos seguintes dados para este trabalho:

Status	Intervalo de Idade	Estado Civil	Cor/Raça	Intervalo de Distância	Forma de Ingresso	Renda per capita familiar	Trabalho	Reserva de Vaga
Concluído	[classificações através de Distribuições de Frequência]	Solteiro	Branca	[classificações através de Distribuições de Frequência]	Ampla Concorrência	Entre 0 e 0,5	Não realiza	Enem/Sisu
Evadido		Casado	Preta		Escola Pública - Renda<=1.5 Salários Mínimos	Entre 0,5 e 1	Realiza	Reingresso Automático
		União Estável	Parda		Escola Pública - Renda>1.5 Salários Mínimos	Entre 1 e 1,5		Enem/Sisu
		Divorciado	Amarela		Reingresso	Entre 1,5 e 2,5		Portador de Diploma - Graduação
		Viúvo	Indígena		Transferência Interna	Entre 2,5 e 3,5		
		Separado				Acima de 3,5		

Figura 7 - Tabela representando dados utilizados para a etapa de mineração de dados. Fonte: autores

Para melhor classificação de dados em questão de processamento, optou-se por transformar as opções de dados em números, como mostra a Figura 8. Estes números são lidos como categorias pelo algoritmo.

status	intervaloIdade	estadoCivil	raca	intervaloDistancia	formalIngresso	RFPC	trabalho	reservaVaga
1	3	1	1	1	1	2	1	1
1	1	1	1	1	1	2	0	2

Figura 8 - Representação parcial dos dados cadastrados no banco de dados utilizado pelo algoritmo. Fonte: autores

A partir da transformação dos dados, foi gerado um *dataset* contendo 27 instâncias salvas no banco de dados, o qual foram submetidos a experimento baseado em mineração de dados com apoio da ferramenta WEKA. Para tanto, foi utilizado o algoritmo RNA *Multilayer Perceptron* [Frank, Hall e Witten, 2016], que teve como objetivo encontrar perfis de alunos propensos à evasão escolar através da tarefa de classificação. Como técnica de validação do experimento, utilizou-se o mecanismo de validação cruzada *K-Fold* [Schreiber et al 2017], com o limite de 10 *folds*.

O resultado apresentou acerto de 18 das 27 instâncias, configurando aproximadamente 66,6% de acurácia.

A partir do experimento realizado, percebeu-se a viabilidade de desenvolvimento de uma ferramenta *web* a fim de assistir gestores do *campus* no delineamento de perfis evasores, descrita na seção 2.3.

2.3 Desenvolvimento da Ferramenta

A última etapa envolvida nesta pesquisa foi o desenvolvimento de um sistema web, cujo objetivo é facilitar a análise e definição de perfis evasores voltado para o gestores do campus. Para isso, o sistema consiste em duas operações principais: cadastro de alunos no banco de dados e predição da evasão a partir dos dados dos alunos cadastrados.

A arquitetura utilizada pode ser dividida em duas camadas principais. A primeira, chamada de *front-end*, corresponde à interação do usuário com o sistema. Esta camada foi implementada utilizando o *framework* Bootstrap 3 [Costa, 2014]. Já o *back-end*, a camada do servidor responsável por fazer operações no banco de dados, foi desenvolvido com a linguagem de programação JAVA [Oracle, 2014], utilizando a tecnologia de *servlets*, pequenos servidores cujo objetivo é receber chamadas HTTP, processá-las e devolver uma resposta ao cliente [Oracle, 2014]. E como banco de dados, foi escolhido o MySQL.

O funcionamento do sistema ocorre da seguinte forma: ao solicitar a opção de *cadastro de aluno*, o sistema apresentará um formulário para que o usuário possa cadastrar os dados do aluno, definidos na seção de mineração de dados (ver Figura 7). Ao enviar o formulário e caso a operação seja bem sucedida, é realizada a atualização no *dataset*, que consiste na atualização das distribuições de frequência (dos intervalos de idade e distância). Esta etapa é necessária pois ao inserir um novo dado, é possível que seus intervalos mudem. A Figura 10 demonstra o diagrama para esta operação.

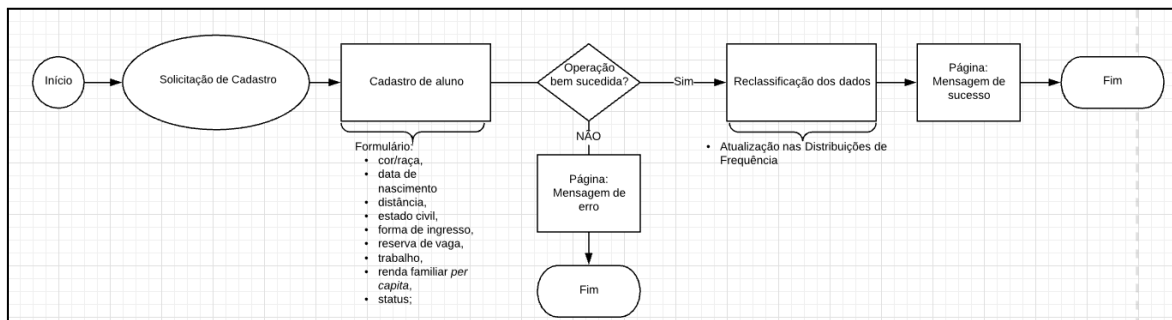


Figura 10 - Diagrama do módulo de cadastro. Fonte: autores.

Já ao solicitar a segunda tarefa permitida pelo sistema, *tarefa de predição*, o sistema apresentará um formulário para que o usuário insira os dados do aluno a ser predito como mostra o diagrama da Figura 9. Ao submeter o formulário, o sistema buscará os dados necessários no banco de dados e treinará o algoritmo de classificação baseado em *RNA Multilayer Perceptron* e, com tal conjunto de treinamento, o sistema informará a probabilidade de evasão do aluno.

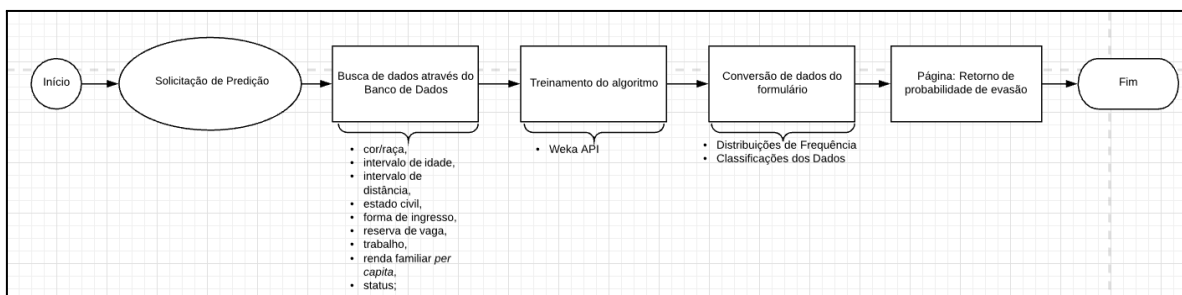


Figura 9 - Diagrama do módulo de predição. Fonte: autores

Para realização da tarefas de classificação e predição a aplicação utiliza a *Application Programming Interface (API)* do software WEKA.

3. Resultados

Com o sistema desenvolvido, o usuário pode interagir com o sistema através de duas formas principais: treinar o algoritmo ou prever a probabilidade de evasão de um aluno. A Figura 11 apresenta a página inicial do sistema:



Figura 11 - Página inicial do sistema APEU. Fonte: autores

A Figura 12 expõe a página de cadastro de aluno através do formulário referente ao documento de requerimento de matrícula. É importante notar que a distância foi determinada como numérica, sendo necessário preencher o número “0” para definir como moradia Panambi, como definido em vermelho abaixo do *input*. Se a operação for inserida no banco de dados, o sistema responderá com uma página de sucesso (Figura 13) caso contrário, uma página de erro será retornada.

Figura 12 - Página de cadastro de usuário. Fonte: autores

Para melhor composição de dados, quando um usuário insere um novo aluno, o sistema executa um algoritmo para realizar a distribuição de frequências das idades, devido a grande variação entre elas [SPSS, sem data].

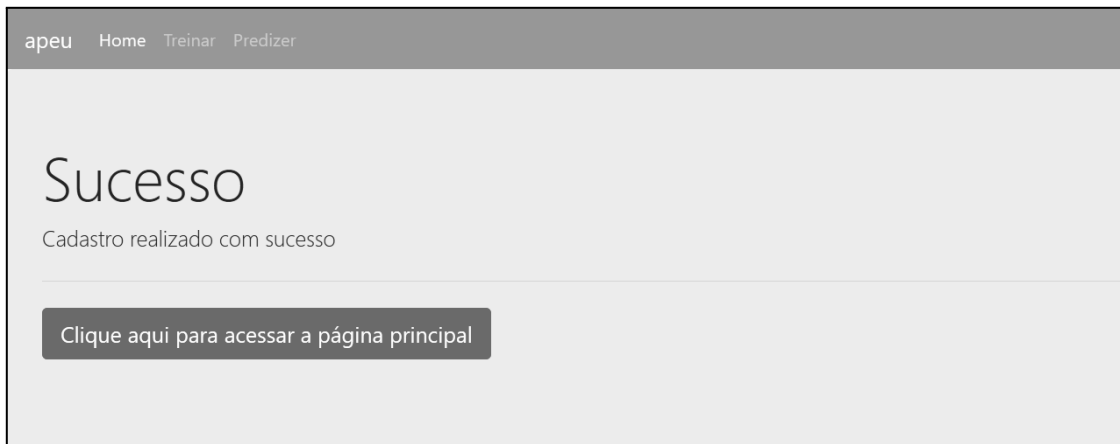


Figura 13 - Página de sucesso. Fonte: autores

A predição do sistema é feita a partir do mesmo formulário da Figura 12 com a exceção de não haver definição de *status*. Ao solicitar uma predição, é feita uma distribuição de frequência de sua idade e distância do município a partir dos dados cadastrados no banco de dados para que seja possível classificar o determinado aluno. Ao preencher o formulário, o sistema retornará a probabilidade de evasão do aluno, como mostra a Figura 14.

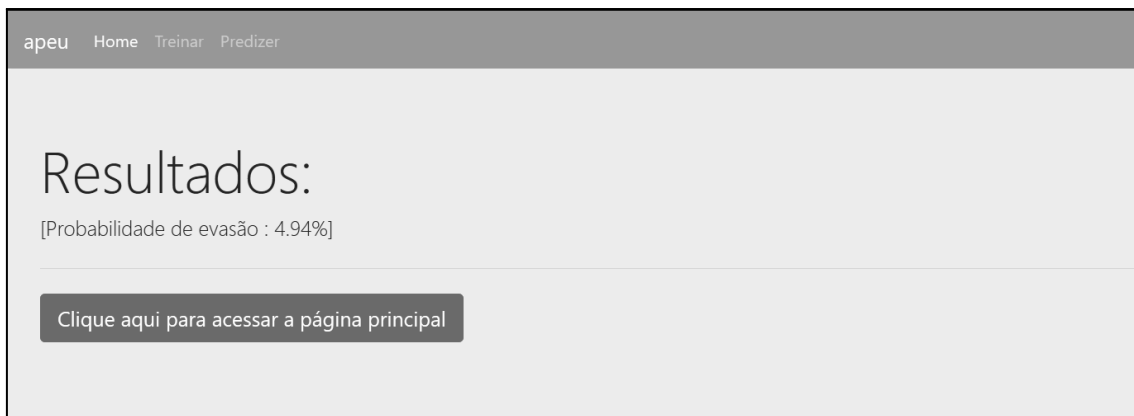


Figura 14 - Página de resultado de predição. Fonte: autores

4. Considerações finais

Esta pesquisa procurou compreender mais a fundo como a Inteligência Artificial pode prever a possibilidade de evasão de alunos ingressantes do IFFAR a partir de seus perfis. Para isso foram desenvolvidos um mapeamento sistemático de literatura, coleta de dados através do método de fichamento, mineração dos dados coletados (KDD) e uma ferramenta para que os gestores do campus possam analisar e diagnosticar o perfil de evasão presente nos dados.

O sistema desenvolvido foi capaz de submeter novos dados e calcular a probabilidade de evasão de alunos. Porém, houve algumas limitações quanto a acurácia do algoritmo devido a pouca quantidade de dados e grande generalização. Pois, apesar de o algoritmo conseguir acertar a maior parte dos conjuntos, ainda assim, os dados cadastrados foram de apenas uma turma com 27 registros que não necessariamente representa o perfil de evasão do *campus* ou ainda do curso. Por exemplo, pode-se notar que todos os alunos que concluíram o curso eram da mesma etnia, o que pode gerar um viés no algoritmo. Portanto, como futuras propostas, sugere-se mais dados e possivelmente um outro estudo acerca de testes com outros algoritmos.

Mas, ainda assim, com mais dados e estudo, o sistema mostra-se capaz de ser aplicado para gestão acadêmica, sendo necessário a construção de um *dashboard* para facilitar a visualização de dados. Também, é possível visualizar o sistema sendo aplicado para mineração de dados em outros contextos, como a predição de outros eventos para fins educacionais. Para facilitar este acesso, o código fonte do projeto pode ser obtido através do *link* no rodapé⁴.

Através da elaboração do assistente, foi possível compreender mais a fundo a utilização da técnica KDD, principalmente quanto a limpeza e transformação de dados para que seja possível priorizar os principais dados para melhor associação. A etapa de mapeamento sistemático, também, fora imprescindível para compreender mais a fundo o estado da arte da questão de pesquisa apresentada, evitando o viés de busca.

Também, no decorrer deste projeto, pôde-se notar a promissoriedade da área de MDE, não apenas pelo potencial de extração de padrões, mas devido a capacidade de análise de resolução de problemas a partir destes padrões.

Como oportunidade para trabalhos futuros, espera-se obter mais dados para que seja possível melhorar a acurácia do sistema proposto, bem como o teste com outros algoritmos de mineração de dados. E por fim, espera-se que o sistema seja utilizado pelos setores acadêmicos.

Agradecimentos

A realização deste projeto foi possível graças à Coordenação de Registros Acadêmicos, pelo fornecimento de dados, à Direção Geral pela permissão de dados, ao *campus* pela confiança e aos professores pelo incentivo.

Referências

Araújo, E. de O. (2018) Sistema de mineração de dados para apoiar a tomada de decisão em uma instituição de ensino superior: o problema da evasão escolar no IFTM, <https://recipp.ipp.pt/bitstream/10400.22/13080/1/Disserta%3%a7%c3%a3o%20-%20Eduardo%20de%20Oliveira%20Araujo.pdf>, Novembro.

Assunção, F. (2012) Estratégias para tratamento de variáveis com dados faltantes durante o desenvolvimento de modelos preditivos, https://teses.usp.br/teses/disponiveis/45/45133/tde-15082012-203206/publico/DissertacaoFernando_vfinal.pdf, Outubro

⁴ <https://github.com/vyk1/APEU>

Baggi, C., Lopes, D. (2011) Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. Avaliação (Campinas), Sorocaba, v. 16, n. 2, p. 355-374, <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-40772011000200007&lng=en&nrm=iso, Junho.

Bezerra, C., Scholz, R., Adeodato, P., Pontes T., Silva, I. (2016) Evasão Escolar: Aplicando Mineração de Dados para Identificar Variáveis Relevantes, <https://www.br-ie.org/pub/index.php/sbie/article/view/6795>, Outubro.

Biolchini, J., Mian, P. G., Natali, A. C. C., Travassos, G. H. (2005) Systematic Review in Software Engineering. <https://www.cos.ufrj.br/uploadfile/es67905.pdf>, Outubro.

Brasil (2008). Lei de N° 11.892, Sobre A Rede Federal De Educação Profissional, Científica E Tecnológica, de Dezembro de 2008, <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Lei/L11892.htm, Abril.

Caelum (sem data) Apostila Java para Desenvolvimento Web <https://www.caelum.com.br/apostila-java-web/servlets/#servlets>, Outubro.

Calixto, K. Segundo, C. Gusmão, R. (2017) Mineração de dados aplicada a educação: um estudo comparativo acerca das características que influenciam a evasão escolar, <http://www.br-ie.org/pub/index.php/sbie/article/view/7674/5469>, Junho.

Candão, J. P., Real, E. M. (2018) A Mineração de Dados Educacionais como Apoio na Análise e Compreensão do Processo de Aprendizagem, <http://www.tise.cl/Volumen14/TISE2018/545.pdf>, Novembro

Coelho, A. (2013) A democratização da universidade pública, https://educere.bruc.com.br/CD2013/pdf/8560_7189.pdf, Junho.

Costa, G. (2014) O que é Bootstrap? <http://www.tutorialwebdesign.com.br/o-que-e-bootstrap/>, Outubro.

Coutinho, E. Bezerra, J. Bezerra, C. I. M. Moreira, L. O. M. (2018) Uma Análise da Evasão em Cursos de Graduação Apoiado por Métricas e Visualização de Dados. <https://www.br-ie.org/pub/index.php/wie/article/view/7869>, Novembro.

Filho, R., Montejunas, P., Hipólito, O., Lobo, M. (2007) A evasão no Ensino Superior Brasileiro, <http://www.scielo.br/pdf/cp/v37n132/a0737132.pdf>, Junho.

Frank, E., Hall, M. A., Witten, I. H. (2016) Weka Workbench 4ª edição.

Han, J., Kamber, M., Pei, J. (2012) 3ª edição. Data Mining. Concepts and Techniques. Novembro.

Kitchenham, B (2004) Procedures for Performing Systematic Reviews, <http://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf>, Outubro

Medeiros, L. B G; Padilha, T. P. P. Mineração De Dados Para Detectar Evasão Escolar Utilizando Algoritmos De Classificação. Ciet: Enped, 2018. <https://cietenped.ufscar.br/submissao/index.php/2018/article/view/623>, Novembro

Nascimento, R. L. S., Junior, G. G. C., Fagundes, R. A. A (2018) Mineração de Dados Educacionais: Um Estudo Sobre Indicadores da Educação em Bases de Dados do INEP, <https://seer.ufrgs.br/renote/article/view/85989>, Outubro.

Oracle. (2014) Java Platform, Enterprise Edition. <https://docs.oracle.com/javaee/7/JEETT.pdf>, Novembro.

Paz, F., Cazella, F. (2017) Identificando o perfil de evasão de alunos de graduação através da Mineração de dados Educacionais: um estudo de caso de uma Universidade Comunitária, <http://www.br-ie.org/pub/index.php/wcbie/article/view/7448/5244>, Junho.

Petersen, K., Feldt, R., Muhtaba, S., & Mattsson, M.,. (2008). Systematic Mapping Studies in Software Engineering. https://www.researchgate.net/publication/228350426_Systematic_Mapping_Studies_in_Software_Engineering, Novembro

Plataforma Nilo Peçanha (2018), <http://resultados.plataformanilopecanha.org/2018/>, Abril.

Plataforma Nilo Peçanha (2019) <http://resultados.plataformanilopecanha.org/2019/>, Abril.

Santos, M. A., Barreto, R. S. (2015) Mapeamento Sistemático, <https://arxiv.org/ftp/arxiv/papers/1504/1504.01027.pdf>, Outubro.

Schreiber, J. N. C., Beskow, A. L., Müller, J. C. T., Nara, E. O. B., Silva, J. I., Reuter, J. W. (2017) Técnicas De Validação De Dados Para Sistemas Inteligentes: Uma Abordagem Do Software SDBAYES, Outubro.

SPSS (sem data) Frequency Distribution - Quick Introduction, <https://www.spss-tutorials.com/frequency-distribution-what-is-it/>, Outubro.