

**DESENVOLVIMENTO DE UM
MODELO DE MINERAÇÃO DE DADOS
PARA PREDIÇÃO DE EVASÃO NA
EDUCAÇÃO A DISTÂNCIA EM SAÚDE:
A EXPERIÊNCIA DA UNA-SUS/UFSJ**

DESENVOLVIMENTO DE UM MODELO DE MINERAÇÃO DE DADOS PARA PREDIÇÃO DE EVASÃO NA EDUCAÇÃO A DISTÂNCIA EM SAÚDE: A EXPERIÊNCIA DA UNA-SUS/ UFSJ

Álisson Oliveira dos Santos; Tarcísio Laerte Gontijo; Caio Godinho Caldeira; Luísa Machado dos Santos Rocha; Marcos Paulo da Cruz Pimenta; Isabela Soares Maia; Viviane Godinho Caldeira

Resumo

A UFSJ/CCO faz parte da Rede UNA-SUS, que pesquisa inteligência artificial em educação a distância, como aprendizado de máquina e mineração de dados para potencialização da experiência educacional. O ambiente EaD (educação a distância) é favorável à obtenção desses dados e à análise de padrões de aprendizagem e trajetória do aluno. Assim, uma técnica de predição de evasão de cursos a distância para profissionais de saúde foi desenvolvida, chamada Rede Neural Perceptron Multicamadas (MLP). A quantidade de registros (número total de alunos avaliados/ matriculados) é dividida em 10 subgrupos. A partir dessa divisão, o método MLP de predição é verificado com os dados existentes e a saída conhecida (aluno evadido ou não) em 9 subgrupos. Em 1 subgrupo, as respostas são testadas (validação). Em resumo, em 9 grupos, sabe-se o resultado final e estimam-se pelo modelo os pesos de cada dado no resultado (evasão ou não). Em 1 grupo, o modelo avalia a resposta e compara com a saída real. Para otimizar a ferramenta, são necessários mais testes e, nesse sentido, trabalhar com outras universidades. Isso poderá potencializar o planejamento da educação profissional em saúde, com vistas a minimizar evasão, garantindo maior engajamento e formação qualificada.

Palavras-chave: Mineração de Dados. Aprendizado Profundo. Educação Continuada.

DEVELOPMENT OF A DATA MINING MODEL FOR PREDICTION OF DROPOUT IN DISTANCE LEARNING IN HEALTH: THE EXPERIENCE OF UNA-SUS UFSJ

Álisson Oliveira dos Santos; Tarcísio Laerte Gontijo; Caio Godinho Caldeira; Luísa Machado dos Santos Rocha; Marcos Paulo da Cruz Pimenta; Isabela Soares Maia; Viviane Godinho Caldeira

Abstract

UFSJ / CCO is part of the UNA-SUS network, which researches artificial intelligence in distance education, such as machine learning and data mining to enhance the educational experience. The distance education environment is favorable to obtain this data and to analyze the student's learning patterns and trajectory. Thus, a technique for predicting the avoidance of distance courses for health professionals was developed, called the multilayer perceptron neural network (MLP). The number of records (total number of students evaluated / enrolled) is divided into 10 subgroups. From this division, the MLP prediction method is verified with the existing data and the known output (student escaped or not) in 9 subgroups. In 1 subgroup, responses are tested (validation). In summary, in 9 groups the final result is known and the weights of each data in the result are estimated by the model (evasion or not). In 1 group, the model evaluates the response and compares it with the actual output. To optimize the tool, more tests are needed and, to that end, work with other universities. This may enhance the planning of professional health education, with a view to minimize dropout, ensuring greater engagement and qualified training.

Keywords: Data mining. Deep Learning. Education. Public Health Professional.

1 INTRODUÇÃO

A Universidade Federal de São João Del-Rei (UFSJ) foi instalada em 21 de abril de 1987 como Fundação de Ensino Superior de São João Del-Rei (FUNREI). No ano de 2002, foi transformada em Universidade. Atualmente, a UFSJ está presente em 4 municípios e seus cursos estão distribuídos em 6 unidades educacionais, sendo 3 em São João Del-Rei, onde está localizada a sede administrativa, 1 em Ouro Branco, 1 em Sete Lagoas e 1 em Divinópolis (Campus Centro-Oeste “Dona Lindu” – CCO). Com a criação de campi em outros municípios, a UFSJ ampliou sua inserção e seu raio de ação. Assim, tem um impacto significativo em Minas Gerais, atuando em 3 mesorregiões (UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI, 2020a).

O Campus Centro-Oeste (CCO) iniciou as atividades em 2008, em Divinópolis. Trata-se do município-polo da região, contando com população estimada de 235.977 mil habitantes e indicadores de saúde ainda insatisfatórios. A UFSJ se insere no território, possibilitando a consolidação da cidade como polo de saúde e formando profissionais graduados nas áreas de Medicina, Enfermagem, Farmácia e Bioquímica. Ademais, são oferecidos 8 programas de pós-graduação, os quais representam a possibilidade de geração de novos conhecimentos e da qualificação dos agentes de saúde da região. Os programas de extensão e pesquisa desenvolvidos no CCO também têm impactado positivamente os sistemas de saúde pública, favorecendo a prevenção e o combate de endemias e epidemias (UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI, 2020b).

Além de suas atividades de ensino presencial, em suas unidades educacionais, a UFSJ oferece cursos na modalidade Educação a Distância (EaD) e desenvolve novos suportes e tecnologias educacionais. Esse cenário foi consolidado em 2008, com a criação do Núcleo de Educação a Distância da UFSJ (NEAD) que culminou com o credenciamento da UFSJ na Universidade Aberta do Brasil (UAB). Atualmente, são oferecidos cursos de graduação, como Administração Pública, Matemática, Pedagogia e Filosofia, além de cursos de pós-graduação lato sensu, mestrado e extensão e aperfeiçoamento, desenvolvidos em 46 Polos de Apoio Presenciais nos estados de Minas Gerais e São Paulo. O número de alunos matriculados nos cursos chega a 7.520 alunos (UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI, 2020c).

Concomitantemente a essa expertise adquirida no âmbito da EaD, a UFSJ/CCO passou a fazer parte da UNA-SUS em 2015, participando da discussão e desenvolvimento da Rede desde então. A UNA-SUS UFSJ tem, entre suas frentes de pesquisa, a inteligência artificial em educação a distância e suas subáreas, como o aprendizado de máquina e mineração de dados para potencialização da experiência educacional.

2 MINERAÇÃO DE DADOS

Nas últimas décadas, muitas das operações, atividades e pesquisas de diversas instituições são armazenadas em grandes bases de dados, muitas vezes subaproveitadas. A técnica da mineração de dados, também conhecida como *Data Mining* (DM), é uma das opções para extrair conhecimento, a partir desses grandes volumes de dados, descobrindo relações ocultas e padrões, gerando regras para prever e correlacionar dados. Dessa forma, a aplicação desse recurso favorece as organizações, por viabilizar a transformação desses arquivos em conhecimento utilizável (CÔRTEZ; PORCARO; LIFSCHITZ, 2002; GALVÃO; DE FÁTIMA MARIN, 2009).

Trata-se de um eficaz processo de cooperação entre homens e máquinas, baseado em técnicas seguramente confiáveis e validadas pela sua expressividade estatística. Nesse sentido, a mineração de dados conta com as seguintes fases: definição do problema; seleção das fontes internas e externas de dados; preparação dos dados, abrangendo o pré-processamento, a reformatação e a análise dos resultados obtidos. Devido à sua relevância prática, esse campo de conhecimento tem se destacado como uma área próspera e de significativa importância, evidenciando-se por meio de congressos científicos (CÔRTEZ; PORCARO; LIFSCHITZ, 2002; GALVÃO; DE FÁTIMA MARIN, 2009).

Essa ferramenta compõe um método maior de pesquisa denominado Busca de Conhecimento em Banco de Dados (Knowledge Discovery in database KDD). No entanto, por ser a etapa em que de fato são empregadas as técnicas de busca de conhecimentos, a DM tornou-se ainda mais conhecida do que o próprio processo de KDD (CÔRTEZ; PORCARO; LIFSCHITZ, 2002; GALVÃO; DE FÁTIMA MARIN, 2009).

O conhecimento obtido por meio da DM tem se mostrado extremamente útil em variadas áreas, dentre elas a medicina. No setor da saúde, pode ser aplicada

na análise da eficácia de tratamentos, na otimização de processos hospitalares, na análise de efeitos colaterais de drogas, dentre outros. Ademais, a DM possibilita previsão de tendências futuras, caracterização melhor do paciente que busca assistência e identificação de terapias médicas de sucesso. Consolida-se, portanto, nesse âmbito, como mecanismo aliado e fundamental para agilizar a busca de conhecimento, a partir de grandes sistemas de informação de saúde e de quaisquer outros bancos expressivos de dados (CÔRTEZ; PORCARO; LIFSCHITZ, 2002; GALVÃO; DE FÁTIMA MARIN, 2009).

No âmbito da educação a distância (EaD), a mineração de dados também tem várias aplicações. Dentre elas, podemos citar a análise de padrões de aprendizagem e trajetória do aluno, o grau de aproveitamento, os motivos que influenciam o baixo desempenho, a frequência e ainda a predição do resultado final. O ambiente EaD é favorável à obtenção dos dados necessários à análise por mineração de dados, tendo em vista que registros textuais em chats, blogs e fóruns de discussões têm excelente potencial para evidenciar as interações aluno-aluno, assim como aluno-professor. Ademais, a coleta de dados permite criar modelos e fazer predições que podem ser aplicadas em outros ambientes virtuais de ensino, inclusive em salas de aula convencionais.

Dessa forma, pode-se afirmar que o uso da mineração de dados tem grande potencial no âmbito educacional, por meio de ações que promovem um ensino eficaz a distância e, nas escolas, por meio do emprego de tecnologias educacionais que complementam o ensino tradicional (BARBOSA; ANDRADE; CARVALHO, 2017; PATRICIO; MAGNONI, 2018).

3 A EVASÃO NO ÂMBITO DA EAD

A evasão na EaD é um problema conhecido e amplamente discutido, sendo considerada, para a maioria dos autores que discutem aprendizagem a distância, a não conclusão do curso por parte do aluno matriculado, em qualquer etapa. (OLIVEIRA; OESTERREICH; ALMEIDA, 2018).

O último Censo EAD.BR, um relatório sobre a aprendizagem a distância no Brasil, conduzido pela Associação Brasileira de Educação a Distância (ABED), apresentou que grande parte dos cursos livres a distância oferecida à população

(14%) tinha uma taxa de evasão que variava entre 26 e 50%. Ou seja, em uma parcela considerável dos cursos a distância, metade dos alunos matriculados não os concluem (ASSOCIAÇÃO BRASILEIRA DE EDUCAÇÃO A DISTÂNCIA, 2019).

Diversos autores têm tentado avaliar e elencar os motivos da elevada evasão. Dentre os motivos elencados, estão os seguintes: a falta de letramento digital, pouco diálogo com as experiências dos estudantes, problemas de comunicação, sobrecarga de trabalho, avaliação, qualidade insuficiente do ensino (BRANCO; CONTE; HABOWSKI, 2020).

Para além da identificação geral dessas barreiras, faz-se necessário individualizar, de forma oportuna, aqueles que estão na iminência de evadir-se de determinada oferta educacional. A partir disso, pode-se tratar de forma personalizada os motivos para tal evasão, estabelecendo planos de ação, no sentido de garantir a continuidade e posterior conclusão do curso.

4 A REDE NEURAL PERCEPTRON

Considerando a incidência da evasão já mencionada no âmbito da EaD, estratégias para identificação tempestiva de alunos com tendência à não conclusão dos cursos que iniciaram, de modo a direcionar uma ação efetiva, parecem factíveis. Nesse sentido e a partir da experiência da UNA-SUS/UFSJ em inteligência artificial, uma técnica de predição de evasão de alunos dos cursos a distância, para profissionais de saúde, foi desenvolvida.

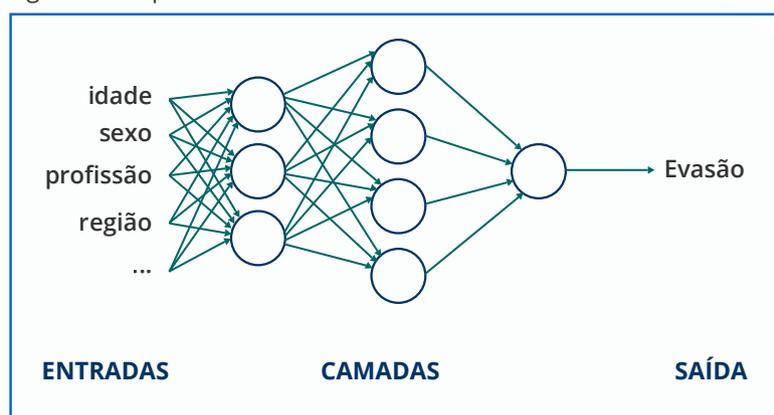
O modelo desenvolvido foi uma rede neural perceptron multicamadas (*Multilayer Perceptron* - MLP). As redes neurais artificiais são modelos computacionais criados para reconhecer padrões, aprender e lidar com diferentes tipos de dados de entrada, gerando uma saída, a qual pode ser única ou múltipla. Essas redes recebem esse nome porque são inspiradas em neurônios de animais (HEIDARI *et al.*, 2019).

Especificamente, o modelo MLP é uma rede neural que envolve diversos dados de entrada, os quais recebem pesos diferentes. A primeira camada (dados e pesos iniciais) é chamada de Camada de Entrada. Os dados dessa camada são interpretados em várias camadas subsequentes (chamadas camadas ocultas), cujo número total vai depender da complexidade dos dados, fato que deu origem ao nome multicamadas. Isso permite uma tomada de decisões sofisticada, em relação

a um modelo de entrada e saída de dados mais simples (de camada única, por exemplo) (CHEN; DO; HSIEH, 2015).

A Figura 1 traz a representação gráfica do modelo MLP proposto pela UNASUS/UFSJ, para predição de evasão de alunos dos cursos a distância voltados para profissionais de saúde. Nele, serão avaliados dados presentes no cadastro do aluno (como idade, sexo, gênero), além de dados relacionados às interações desse aluno com o curso (número de acessos, interações com a plataforma, notas).

Figura 1 - Esquema do MLP desenvolvido



Fonte: Elaboração dos autores (2020).

Como observado, alguns dados devem ser colhidos em tempo real, além de serem atualizados constantemente. Isso permite uma maior fidelidade das informações e uma saída mais precisa. Conforme já mencionado, o MLP retornará uma saída binária, indicando se o aluno avaliado está ou não em risco de evasão. O Quadro 1 apresenta o conjunto total de dados que serão avaliados.

Quadro 1 - Total de dados de entrada que serão avaliados pelo MLP

Dados de cadastro		Dados de interação
Sexo	Escolaridade	Número de acessos
Gênero	Estado Civil	Número de interações
Idade	Região	Notas obtidas até o momento
Raça/Cor	Vínculo com Equipe de Saúde da Família	Trimestre da oferta
Profissão	-	-

Fonte: Elaboração dos autores (2020).

O algoritmo foi desenvolvido utilizando a linguagem de programação Python e a biblioteca TensorFlow, a qual é livre e de código aberto. A validação do modelo se encontra em desenvolvimento e utilizará o método *k-fold* de validação cruzada (JUNG, 2018). Nesse método, a quantidade de registros (no caso específico, o número total de alunos avaliados/matriculados) é dividida em 10 subgrupos. A partir dessa divisão, 9 subgrupos se tornam grupos de treinamento, ou seja, o método MLP de predição é verificado com os dados existentes e a saída conhecida (aluno evadido ou não). Em 1 subgrupo, as respostas são testadas (validação). Em resumo, em 9 grupos se sabe o resultado final e estima-se pelo modelo os pesos de cada dado no resultado (evasão ou não). Em 1 grupo, o modelo avalia a resposta e compara com a saída real. Isso ocorre de forma cíclica, com todos os 10 subgrupos, sendo 1 o subgrupo de teste a cada rodada.

A Figura 2 mostra uma representação de como o método de validação funcionará para a avaliação da acurácia do modelo.

Figura 2 - Representação do método *k-fold* de validação cruzada



Fonte: Elaboração dos autores (2020).

5 CONSIDERAÇÕES FINAIS

O modelo MLP sugerido e o método de validação ainda estão em desenvolvimento e constante evolução. Para otimizar a ferramenta, são necessários mais testes e, nesse sentido, o trabalho com outras universidades que oferecem cursos diversos, em diferentes contextos.

Para além da evolução do MLP, novos horizontes de discussão sobre motivações para ingresso e permanência de alunos, no cenário da EaD, poderão

ser levantados. Isso poderá potencializar o planejamento da educação profissional permanente e a EaD em saúde, com vistas a minimizar a evasão, garantindo maior engajamento e formação qualificada dos ingressantes nos cursos.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE EDUCAÇÃO A DISTÂNCIA - ABED. **Censo EAD.BR 2018:** relatório analítico da aprendizagem a distância no Brasil. Curitiba: Intersaberes, 2019. Disponível em: http://abed.org.br/arquivos/CENSO_DIGITAL_EAD_2018_PORTUGUES.pdf. Acesso em: 02 out. 2020.

BARBOSA, A. A. S.; ANDRADE, F. S.; CARVALHO, R. N. Mineração de dados em ambientes virtuais de aprendizagem: Aportes para a pesquisa em educação a distância. **Interfaces Científicas - Educação**, Aracaju, v.6, n.1, p. 125 - 136, out. 2017. Disponível em: <https://periodicos.set.edu.br/index.php/educacao/article/view/4347/0>. Acesso em: 29 jul. 2020.

BRANCO, L. S. A.; CONTE, E.; HABOWSKI, A. C. Evasão na educação a distância: pontos e contrapontos à problemática. **Avaliação (Campinas)**, Sorocaba, v. 25, n. 1, p. 132-154, abr. 2020. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-40772020000100132&lng=en&nrm=iso. Acesso em: 27 jul. 2020.

CHEN, J-F.; DO, Q. H.; HSIEH, H-N. Training Artificial Neural Networks by a Hybrid PSO-CS Algorithm. **Algorithms**, v. 8, p. 292-308, 2015. Disponível em: <https://www.mdpi.com/1999-4893/8/2/292>. Acesso em: 27 jul. 2020.

CÔRTEZ, S. D. C.; PORCARO, R. M.; LIFSCHITZ, S. Mineração de Dados – Funcionalidades, Técnicas e Abordagens. **PUC-Rio Informática**, p. 35, maio 2002. Disponível em: ftp://obaluae.inf.puc-rio.br/pub/docs/techreports/02_10_cortes.pdf. Acesso em: 27 jul. 2020.

GALVÃO, N. D.; DE FÁTIMA MARIN, H. Técnica de mineração de dados: uma revisão da literatura. **ACTA Paulista de Enfermagem**, v. 22, n. 5, p. 686-690, 2009.

HEIDARI, A. A. et al. An efficient hybrid multilayer perceptron neural network with grasshopper optimization. **Soft Computing**, v. 23, n. 17, p. 7941–7958, 2019. Disponível em: <https://link.springer.com/article/10.1007/s00500-018-3424-2>. Acesso em: 27 jul. 2020.

JUNG, Y. Multiple predicting K-fold cross-validation for model selection. **Journal of Nonparametric Statistics**, v. 30, n. 1, p. 197–215, 2018. Disponível em: <https://amstat.tandfonline.com/doi/abs/10.1080/10485252.2017.1404598#.XyHNXihKjIU>. Acesso em: 27 jul. 2020.

OLIVEIRA, P. R. de; OESTERREICH, S. A.; ALMEIDA, V. L. de. Evasão na pós-graduação a distância: evidências de um estudo no interior do Brasil. **Educ. Pesqui.**, São Paulo, v. 44, p. 1-20, 2018. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-97022018000100307&lng=pt&nrm=iso. Acesso em: 27 jul. 2020.

PATRICIO, T.; MAGNONI, M. da G. Mineração de dados e big data na educação. **Revista Geminis**, v. 9, n. 1, p. 57-75, 22 jun. 2018. Disponível em: <http://www.revistageminis.ufscar.br/index.php/geminis/article/view/365>. Acesso em: 29 jul. 2020.

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI - UFSJ. **A UFSJ**. Disponível em: https://ufsj.edu.br/dplag/a_ufsj.php. Acesso em: 25 jul. 2020a.

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI - UFSJ. **O Campus Centro-Oeste Dona Lindu**. Disponível em: <https://ufsj.edu.br/cco/index.php>. Acesso em: 25 jul. 2020b.

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI - UFSJ. **Sobre o NEAD**. Disponível em: <https://www.nead.ufsj.edu.br/portal/nead-info-bar-about/nead-info-bar-about-outros>. Acesso em: 25 jul. 2020c.

AUTORES



Álisson Oliveira dos Santos

Médico de Família e Comunidade; Mestre em Telemedicina e Telessaúde; Doutorando em Ciências da Saúde; Professor da Universidade Federal de São João Del-Rei - Campus Centro-Oeste (UFSJ/CCO); Coordenador Adjunto da UNA-SUS UFSJ; Pesquisador colaborador da Fiocruz/Brasília.



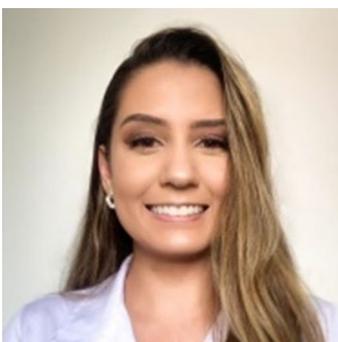
Tarcísio Laerte Gontijo

Enfermeiro; Especialista em Gestão de Sistemas e Serviços de Saúde; Mestre em Enfermagem; Doutor em Ciências da Saúde; Professor da Universidade Federal de São João Del-Rei - Campus Centro-Oeste (UFSJ/CCO); Coordenador da UNA-SUS UFSJ.



Tarcísio Laerte Gontijo

Acadêmico de Medicina da Universidade Federal de São João Del-Rei - Campus Centro-Oeste (UFSJ/CCO).



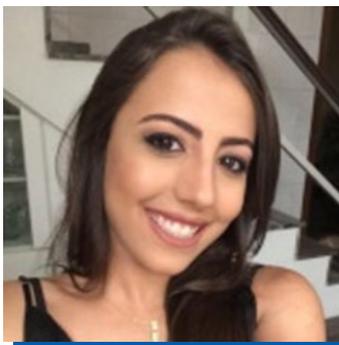
Luísa Machado dos Santos Rocha

Acadêmica de Medicina da Universidade Federal de São João Del-Rei - Campus Centro-Oeste (UFSJ/CCO).



Marcos Paulo da Cruz Pimenta

Acadêmico de Medicina da Universidade Federal de São João Del-Rei - Campus Centro-Oeste (UFSJ/CCO).



Isabela Soares Maia

Acadêmica de Medicina da Universidade de Itaúna (UIT).



Viviane Godinho Caldeira

Acadêmica de Medicina da Faculdade de Minas (FAMINAS-BH).