**PREVISÃO DE ESTUDANTES COM RISCO DE EVASÃO UTILIZANDO TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL**

Amanda Caetano Amorim, Marcia Ferreira Cristaldo, Leandro de Jesus

Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso do Sul – Aquidauana-MS

amanda.amorim@estudante.ifms.edu.br, {marcia.cristaldo, leandro.jesus}@ifms.edu.br

Área/Subárea: Ciências Exatas e da Terra Tipo de Pesquisa: Cientifica

**Palavras-chave:** Mineração de Dados Educacionais, Criação de Atributos, Evasão.

**Introdução**

Os estudos ligados à evasão iniciaram com teorias que buscam explicar a evasão e retenção. Apesar de não existir um conceito coeso, autores como Tinto (1975) abordam o modelo de integração do estudante, frisando que a decisão de evadir surge em função da falta de adaptação com o meio acadêmico/social da instituição, influenciada pelas características individuais e expectativas para a carreira ou curso e intenções e assumidas antes do início do curso. A partir desta premissa, este projeto investigou quais são os fatores intrínsecos e quais os extrínsecos à escola que contribuem para a evasão de alunos na educação profissional, um dos resultados importante verificado foi que a escolaridade da família e a rede mundial de computadores tem papéis de extrema importância na decisão de evasão, observando também que a maior taxa evasão está ligada aos alunos que estão nos primeiros períodos.

**Metodologia**

Com base nos objetivos a serem conquistados, essa pesquisa pode ser definida como exploratória. Segundo Gil (2010), a pesquisa exploratória visa possibilitar maior proximidade com o problema proposto, tornando-o mais explícito. A metodologia segue os passos: **1)** Mineração de Dados: A KDD (*knowledge-discovery in databases*), ou extração de conhecimento no português, é um processo de extração de informações de base de dados. Na base de dados usada neste trabalho temos a descoberta supervisionada, sendo composto pelos dados de todos os estudantes desde **2014.1. 2)** MLP: Perceptron Multicamadas (MLP — Multi Layer Perceptron) é uma rede neural com uma ou mais camadas ocultas com um número indeterminado de neurônios. A técnica MLP possui vantagens na sua arquitetura, a qual a diferenciação entre uma rede e outra fica por conta do número de neurônios da camada oculta (ABDI *et al*., 1999). **3)** Processamento à Aplicação do Algoritmo de RNA: O processo de KDD foi utilizado então os dados do questionário socioeconômico dos estudantes fornecido pela coordenação de Permanência e êxito do IFMS. O vetor de entrada da rede neural artificial - RNA é composto por 54 parâmetros na camada de entrada e a saída da rede é constituída pela classe chamada “situação”. Os dados da situação dos alunos foram fornecidos pela coordenação do curso TII (Técnico Integrado em Informática) por meio de requerimento na secretaria acadêmica do Câmpus. Para criar o banco de dados na etapa de mineração, deve-se criar arquivos do tipo ARFF. **4)** *Software Weka:* Segundo Weka (2020) o *software* é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Ele contém ferramentas para preparação de dados, classificação, regressão, *clustering*, mineração de regras de associação e visualização. **5)** Configuração dos Algoritmos: Nessa etapa foram utilizados 54 atributos iniciais, no qual o algoritmos CFS selecionou 8 atributos, sendo: estado civil do estudante, sexo dos responsáveis, renda familiar, escolaridade dos responsáveis, transporte utilizado, utilização de internet, período matriculado. Após selecionado os melhores atributos, foi configurado o algoritmo de classificação MLP, após a conclusão do teste usa-se um conjunto diferente de 9 partes para treino e 1 para teste até que todas as 10 partes tenham sido usadas para testes.

**Resultados e Análise**

Na Tabela 1 pode-se verificar os resultados dos algoritmos MLP mostrando sua principal métrica, a curva de ROC, um dos valores mais importantes produzidos pela Weka, no qual mostra uma ótima classificação com valores próximos de 1. Outra análise é da Taxa de TP Rate: taxa de verdadeiros positivos (instâncias classificadas corretamente como uma determinada classe), sendo seu valor próximo de 1. Em relação às tarefas de classificação é importante o número de instâncias corretamente classificadas, porém a estatística de Kappa a qual indica o quão concordante (e também coeso) aquele dado está classificado dentro da tarefa de classificação, tendo como resultado de 70% de confiabilidade, e sua acurácia de 86%. A MAE apresenta taxa de erro usada na previsão próxima de 0.

**Tabela 1.** Resultados do algoritmo MLP.

# **Fonte.** Autoria própria.

|  |  |
| --- | --- |
| Métricas | Acurácia |
| Acurácia | 86% |
| Área Média ROC | 0,857 |
| Média TP Rate  | 0,861 |
| MAE | 0,1594 |
| Kappa | 0,7072 |

# Analisando a matriz de confusão na Tabela 2 da MLP.

# TABELA 2 - Matriz de confusão

|  |  |  |
| --- | --- | --- |
| a | b | Classificação |
| 217 | 26 | a= não evadido |
| 29 | 125 | b= evadido |

# Fonte: Autoria própria.

# As instâncias classificadas corretas e incorretamente mostram a porcentagem de instâncias de teste que foram classificadas de forma correta ou não. Assim, seus números brutos são mostrados na matriz de confusão (Tabela 2), com a e b representando os rótulos de classe. Aqui houve 397 instâncias, então as porcentagens e os números brutos se somam, aa + bb = 217 + 125 = 342, ab + ba = 26 + 29 = 55. Na matriz verifica-se que 125 estudantes foram classificados corretamente como evadidos, 29 foram classificados não evadidos. Já os não evadidos 217 estudantes foram classificados corretamente, 26 foram classificados como evadidos. A porcentagem de instâncias classificadas corretamente é frequentemente chamada de exatidão ou precisão de amostra. Ele tem algumas desvantagens como uma estimativa de desempenho por isso deve-se avaliar a curva de ROC e a medida de Kappa mostrada acima.

1. Discussões

A autora Narciso (2015) um dos trabalhos mais recentes sobre evasão nos Instituto Federais (IFs), revelou que os fatores que mais contribuíram para a evasão foram: enorme dificuldade nas disciplinas técnicas e exatas (82%); atividades fora dos IFs que atrapalhavam a permanência na escola (74%) e dentre essas atividades o “trabalho” foi o que mais contribuiu com a evasão (61,53%); falta às aulas (60%), sendo que os motivos das faltas às aulas eram “notas baixas” e “problemas com a instituição” (61,53%).

Em contraponto, este projeto investigou quais são os fatores intrínsecos e quais os extrínsecos à escola que contribuem para a evasão de alunos na educação profissional no campus Aquidauana, revelando que a escolaridade da família, rede mundial de computadores e cursar os primeiros semestres têm papéis de extrema importância na decisão de evasão.

**Considerações Finais**

A evasão dos cursos técnicos dos Institutos Federais é um fenômeno em crescimento e tornou-se foco de preocupação para gestores educacionais. Com intuito de minimizar os gastos com a evasão, o projeto utilizou dados do questionário socioeconômico para apoiar a tomada de decisão no uso de técnicas de mineração de dados. O algoritmo MLP mostrou-se relevante para previsão do estudante com instâncias classificadas corretamente de 86%.

Com a abordagem computacional proposta para a análise da evasão, permite-se identificar padrões que podem auxiliar a tomada de decisão dos gestores educacionais, utilizando-se indicadores e/ou conjunto de regras que permitem avaliar a possibilidade da evasão de cada aluno, abrangendo do nível estratégico até o nível operacional (coordenações de curso), com o uso de técnicas de mineração de dados.

**Referências**

ABDI, H.; VALENTIN, D.; EDELMAN, B. “Neural networks. Sage University Papers Series on Quantitative Applications in the Social Sciences”, 07-124. Thousands Oaks, CA, 1999.

GIL, A. C.. “**Como elaborar projetos de pesquisa**”. 5ª Ed. São Paulo: Atlas, 2010.

NARCISO, L. G. S. “**Análise da** evasão **nos cursos técnicos do instituto federal do norte de Minas Gerais – Câmpus Arinos: exclusão da escola ou exclusão na escola?**”. 2015. 262. Dissertação (Mestrado em Sociologia Política) - Universidade Federal de Santa Catarina. Universidade Federal de Santa Catarina, Florianópolis.

TINTO, V. “Dropout from higher education: A Theoretical syntesis of resent research. Review educational research”, p. 89-125, 1975.

*WEKA.* Software de Mineração de Dados. Disponívele em:<[Weka 3 - Data Mining with Open Source Machine Learning Software in Java (waikato.ac.nz)](https://www.cs.waikato.ac.nz/ml/weka/) >. Acesso em: 2 de março de 2020.