

Predição de evasão de estudantes *non-majors* em disciplina de introdução à programação

André Fabiano Santos Pereira, Leandro Silva Galvão de Carvalho, Eduardo James Pereira Souto

Instituto de Computação – Universidade Federal do Amazonas (UFAM)
Av. General Rodrigo Octávio, 6200 – Coroado I – Manaus – AM – Brasil

{andre.fabiano, galvao, esouto}@icomput.ufam.edu.br

Resumo. *Evasões de alunos são comumente observadas em turmas introdutórias de programação de cursos de graduação de ciências exatas e de engenharia (CS1 para non-majors). Mediante a aplicação de técnicas de mineração de dados sobre informações de estudantes de uma instituição de ensino superior brasileira, observou-se que elementos socioeconômicos, alheios ao ambiente de ensino-aprendizagem, podem atuar tanto como catalisadores quanto neutralizadores desse processo de evasão. Baseado nessa inferência e no uso de algoritmos de aprendizagem de máquina supervisionada, este trabalho subsidia a elaboração de um modelo preditivo para identificação precoce de estudantes non-majors com risco de evasão em CS1.*

Abstract. *Student dropout is commonly observed in introductory programming courses at undergraduate courses in exact sciences and engineering (CS1 for non-majors). Through the use of data mining techniques on student information from a Brazilian higher education institution, it was observed that socioeconomic elements, unrelated to the teaching-learning environment, can act both as catalysts and as neutralizers of this dropout process. Based on this inference and the use of supervised machine learning algorithms, this work supports the elaboration of a predictive model for early identification of dropout non-majors students at CS1.*

1. Introdução

A evasão estudantil em disciplinas introdutórias de programação oferecidas em cursos superiores das áreas de ciências exatas e de engenharia, conhecidas na literatura como *Computer Science 1 (CS1)*, constitui-se em um desafio contínuo enfrentado por instituições de ensino [Digiampietri *et al.* 2016]. Estudantes dessas áreas costumam apresentar maior dificuldade no transcurso da disciplina [Santana *et al.* 2017], produzindo quadros não raros de reprovação ou evasão nessas matérias.

Ademais, a crescente importância e relevância da tecnologia em profissões distintas da área computacional (humanas, ciências exatas e engenharias) impôs a necessidade de serem oferecidas disciplinas como as de introdução à programação inseridas nos mais variados currículos de cursos de graduação [Dawson *et al.* 2018]. Estes casos são conhecidos como CS1 para *non-majors*.

Técnicas de mineração de dados educacionais (*Educational Data Mining, EDM*) têm sido aplicadas para construção de modelos preditivos de evasão em CS1 [Quille *et*

al. 2017] [Pereira *et al.* 2019] [Petersen *et al.* 2016] visando à condução de iniciativas institucionais e pedagógicas mais eficientes de combate à evasão [Júnior *et al.* 2017].

Não obstante, a mineração de dados mostrou-se viável para solucionar diversos problemas relacionados à investigação de informações em bases de dados [Manhães *et al.* 2012]. Este trabalho dará ênfase à utilização da mineração de dados dentro do contexto educacional (*educational data mining*) para investigação e extração de informações relevantes de estudantes [Brito *et al.* 2015] de cursos de graduação das áreas de exatas e engenharias da Universidade Federal do Amazonas (UFAM) a fim de identificar os principais fatores comuns aos grupos daqueles que evadem de CS1.

Nesse contexto, este estudo busca subsidiar a elaboração de um modelo de predição de evasão em CS1 baseado na utilização de atributos sociodemográficos. Visa, ainda, apresentar o resultado de uma análise exploratória baseada nas seguintes questões de pesquisa: **(Q1)** Quais dos atributos coletados a partir de informações socioeconômicas e do cadastro acadêmico dos alunos têm maior correlação com a evasão de estudantes em CS1? **(Q2)** Quais os vínculos entre camadas distintas de dados (sociais, econômicos, familiares e de conceitos prévios) de atributos com a evasão de estudantes em CS1?

O diferencial deste trabalho em curso em relação à grande maioria dos demais existentes reside primeiramente no fato deste concentrar-se no problema da evasão em CS1 para *non-majors* em meio a maioria absoluta de pesquisas que trata sobre evasão em CS1 para alunos em cursos de graduação da área de tecnologia [Petersen *et al.* 2016] ou sobre evasão do ambiente acadêmico [Júnior *et al.* 2017].

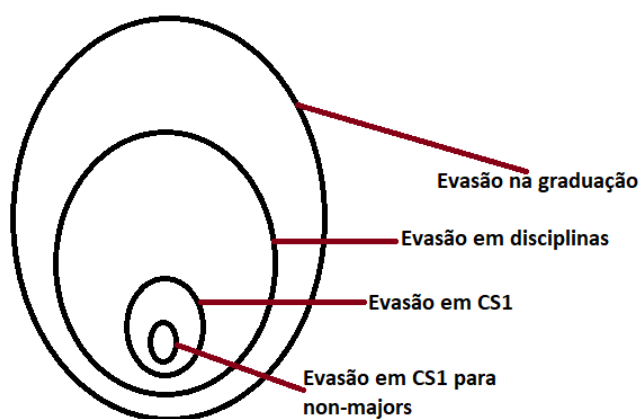


Figura 1: Ilustração da distribuição quantitativa de trabalhos por tipo de evasão. [Fonte: Própria]

Este artigo está estruturado da seguinte maneira: a Seção 2 descreve os trabalhos relacionados. A Seção 3 relata a metodologia de avaliação aplicada. A Seção 4 apresenta os dados e o contexto educacional da pesquisa. A Seção 5 discute os experimentos e os resultados. Por fim, a Seção 7 apresenta as conclusões e trabalhos futuros.

2. Trabalhos relacionados

A evasão é um fenômeno complexo, associado com a não concretização de expectativas e reflexo de múltiplas causas que precisam ser compreendidas no contexto socioeconômico, político e cultural, no sistema educacional e nas instituições de ensino. Caracteriza-se por ser um processo de exclusão determinado por fatores e variáveis internas e externas às instituições de ensino [Fritsch 2015].

A identificação de fatores que possam contribuir para antecipar a evasão em CS1 tem sido objeto de interesse em diversas pesquisas em informática na educação [Petersen *et al.* 2016] [Quille *et al.* 2017], haja vista que a observação e o combate a sinais anteriores à evasão acadêmica em cursos de graduação, como o abandono em CS1, podem resultar em uma maior chance de sucesso no processo manutenção de estudantes no ambiente universitário [Brito *et al.* 2015].

Como complemento a essa visão, Digiampietri *et al.* (2016) propõem um modelo baseado no classificador *Rotation Forest* para identificação do aluno com elevado risco de evasão com base nos seu histórico escolar nas disciplinas do primeiro ano do curso de graduação. Os autores concluem que caso o estudante não perceba a relevância de uma disciplina ou qualquer outra atividade exigida em sua formação e/ou atuação profissional, ele facilmente poderá perder o interesse pela disciplina, incorrendo em reprovações e evasões.

Brito *et al.* (2015) propõem modelos baseados nos algoritmos *NaiveBayes*, *J48* e *AdaBoostM1*, para a identificação de estudantes com risco de evasão a partir da observação das notas obtidas pelos estudantes na prova de ingresso na instituição e nas disciplinas do primeiro semestre do curso de graduação.

Giraffa *et al.* (2015) propõem o uso de um questionário para inferir as possíveis causas relacionadas ao cancelamento da disciplina por parte de alunos. O questionário foi aplicado em alunos que cancelaram a inscrição na disciplina pelo menos uma vez nos três semestres analisados. Os autores concluíram que 74% dos alunos declararam que o critério “falta de tempo para estudar” foi um fator importante ou decisivo para o cancelamento da disciplina e que para 21% dos alunos desistentes a ausência de conhecimento prévio de raciocínio-lógico matemático oriundo do ensino fundamental e médio foi fator decisivo para a evasão da disciplina.

Por outro lado, antigos estereótipos vinculados ao sexo do estudante também já foram desqualificados em estudos como o de Quille *et al.* 2017, demonstrando estatisticamente que homens e mulheres não diferem na capacidade de permanência na disciplina.

O trabalho em curso difere dos demais analisados por nem utilizar dados do histórico escolar nem por aplicar questionários em fase posterior à evasão desses estudantes, buscando tão somente análise sobre dados socioeconômicos de tais alunos na construção de um modelo preditivo de evasão com uso de algoritmos de aprendizagem de máquina.

3. Metodologia Aplicada

No trabalho corrente, técnicas de mineração de dados educacionais foram aplicadas em informações provenientes de duas bases de dados educacionais: histórico de avaliações com resultado final de um sistema de juiz *online* e registros de discentes no sistema de controle acadêmico. Em seguida, foram definidos os atributos que seriam extraídos, conforme Tabela 1, limpeza e transformação dos dados selecionados, extração dos dados da camada de interesse e avaliação dos dados minerados a fim de gerar o conhecimento sobre os mesmos.

Os atributos utilizados na pesquisa foram extraídos de um questionário socioeconômico preenchido nas primeiras aulas de CS1 num juiz *online* de uma instituição de ensino superior brasileira, chamado de *Codebench*

(<http://codebench.icomp.ufam.edu.br>), utilizado na implementação de uma metodologia híbrida de ensino e aprendizagem, mesclando o ensino presencial com atividades *online* baseadas em um ambiente virtual de aprendizagem (AVA) [Wasik *et al.* 2018].

Os atributos coletados (Tabela 1) foram comparados ao atributo de saída correspondente ao Resultado de cada aluno em *CSI*: evasão ou não evasão. As informações de interesse, extraídas de dados oriundos do *Codebench*, para cada ex-aluno com registro de evasão na disciplina *CSI*, e consideradas pelo presente trabalho estão descritas na Tabela 1.

Tabela 1: Variáveis consideradas durante a pesquisa [Fonte própria]

Item	Atributo	Valores
1	A01_sexo	Masculino ou feminino
2	A02_estado_civil	Solteiro, casado, divorciado e viúvo
3	A03_experiencia_programacao	Experiência em alguma linguagem de programação (sim ou não)
4	A04_faixa_etaria	16 a 18 anos – 19 a 22 anos – 23 a 27 anos – 28 a 32 anos – 33 a 39 anos – 40 a 49 anos – acima de 50 anos
5	A05_filhos	Possuir ou não filhos
6	A06_origem_ensino_medio	Público convencional – privado convencional – médio técnico
7	A07_vaga_acao_afirmativa	Ampla concorrência - Cota Independente de renda - Cota com renda baixa – Outros
8	A08_experiencia_trabalho	Possui experiência laboral (sim ou não)
9	A09_curso	Um dos 11 cursos de ciências exatas e engenharia analisados
10	A10_grupo_curso	Ciências Exatas ou Engenharias
11	A11_acesso_internet	Possui acesso à internet (sim ou não)
12	A12_pc_casa	Computador em casa (sim ou não)
13	A13_turno_ensino_medio	Matutino - Vespertino - Noturno - Integral

Na obtenção de dados brutos, foram extraídos registros de resultados finais em *CSI* de alunos e ex-alunos de cursos de graduação das áreas de engenharia e ciências exatas, disponíveis na base de dados do sistema de gestão educacional da instituição. Foram extraídos, ainda, dados brutos de informações socioeconômicas dos alunos a partir da base de dados do *Codebench*. Os históricos de resultados finais em *CSI* extraídos estão vinculados a estudantes ativos e egressos do quadro de discentes da universidade.

No presente trabalho, tratou-se a predição como um problema de classificação binária [Ramos *et al.* 2018], no qual, dadas as informações do histórico de um aluno, o classificador tenta identificar se esse aluno irá ou não se evadir da disciplina. Para isso, foram realizados experimentos de seleção de atributos para identificar quais as características mais relevantes na predição de evasão. Além disso, visando à generalização do modelo a ser obtido, classificadores foram treinados utilizando a base de dados coletada particionada em 66% para treino e 34% para teste.

Os resultados esperados para este estudo são: o aluno conseguiu concluir a disciplina (independentemente de ser aprovado ou não) ou o aluno evadiu-se da mesma.

Como ferramenta de aprendizagem de máquina, foi utilizado o pacote de software *Weka* [Hallet *et al.* 2009], sistema pelo qual foram utilizados os algoritmos *CfsSubsetEval* para seleção de atributos e o *C4.5* para ranqueamento dos atributos mais relevantes [Witten *et al.* 2011], descritos na Tabela 4. Em seguida, modelos de

classificação foram treinados utilizando diversos algoritmos de aprendizagem de máquina supervisionada como *Neural Network*, *SVM*, *AdaBoost*, *KNN* e *Random Forest*. Por fim, os modelos foram avaliados e comparados com base em sua taxa de acerto e os resultados apontaram a possibilidade de predição de estudantes com risco de evasão ainda nas semanas iniciais dos cursos, com taxas de desempenho próximas a 73% na maioria dos cenários.

4. Descrição dos dados coletados e do contexto educacional

Neste estudo, foram coletados dados anonimizados de estudantes de 11 cursos de graduação presencial nas áreas de engenharia e ciências exatas em que a disciplina de Introdução à Programação (CS1) era um componente curricular obrigatório. Os dados são provenientes de turmas realizadas nos primeiros semestres letivos dos anos de 2016, 2017 e 2018 totalizando 1.040 registros de matrícula, entre os quais 872 correspondiam a alunos únicos vinculados aos dados da base do sistema de gestão da instituição. A Tabela 2 apresenta uma síntese dos dados coletados durante a pesquisa.

Tabela 2: Síntese de informações sobre o *dataset* utilizado [Fonte própria]

Item	Características
População-alvo	Alunos de graduação das áreas de engenharia e ciências exatas
Limitação de escopo	Disciplina introdutória de programação (<i>CS1</i>)
Fonte de dados	Juiz <i>online</i> e sistema de controle acadêmico
Períodos letivos avaliados	2016/1, 2017/1 e 2018/1
Quantidade total de registros de matrícula	1.016
Cursos de graduação	11 cursos de ciências exatas e engenharia (<i>non-majors</i>)
Atributos selecionados	treze atributos
Variável dependente	Situação final na disciplina (evasão ou não evasão)

Do total de alunos, 64,9% eram homens e 35,1% eram mulheres. Além disso, 69,8% não possuíam experiência anterior em programação e apenas 64,4% do total era oriundo de escola pública convencional. A média geral histórica de evasão em *CS1* na instituição de ensino superior utilizada, no período pesquisado, girou em torno de 37,0%, sendo que o curso com maior evasão foi o de Licenciatura em Matemática Matutino com 61,8% de evasão, seguido de Bacharelado em Matemática Diurno (licenciatura) e Licenciatura em Matemática Noturno. Em contrapartida, o curso de Engenharia de Produção, apresentou o menor número de evasão entre todos, conforme pode ser verificado na Tabela 3.

Tabela 3: Índice médio de evasão em CS1 por curso [Fonte própria]

Curso de Graduação	% Evasões	% Aprovações	% Reprovações
Matemática Licenciatura Matutino	61,8	31,9	6,30
Matemática Bacharelado Diurno	58,8	34,9	6,30
Matemática Licenciatura Noturno	55,4	43,4	1,20
Física Licenciatura Noturno	54,2	35,8	10,00
Física Bacharelado Noturno	52,8	25,4	21,80
Engenharia de Materiais	34,0	44,0	22,00
Estatística	33,3	60,2	6,50

Engenharia Química	27,9	62,8	9,30
Engenharia de Petróleo e Gás	22,8	67,6	9,60
Engenharia Mecânica	22,8	68,4	8,80
Engenharia de Produção	22,2	70,4	7,4

5. Análise e discussão sobre os resultados

O modelo preditivo de evasão dos estudantes em *CSI* alcançou 70,1% de acurácia com uso do classificador *AdaBoost* no conjunto de testes. Destaca-se que, na tentativa de identificar modelos com acurácia superior ao encontrado, foram avaliados outros classificadores baseados em algoritmos de aprendizagem de máquina supervisionada. Entretanto, nenhum obteve resultados superiores ao primeiro, conforme detalhado na Tabela 4.

Tabela 4: Resultados de acurácia e precisão por modelo [Fonte própria]

Método	Acurácia	Precisão	Curva ROC
AdaBoost	0,701	0,701	0,701
Random Forest	0,679	0,676	0,720
Neural Network	0,669	0,679	0,701
Naive Bayes	0,663	0,663	0,703
KNN	0,640	0,630	0,653
SVM	0,608	0,080	0,622

A figura 1 exibe a área sobre a curva ROC (*Receiver Operating Characteristics*) utilizada para gerar uma representação gráfica ilustrando o desempenho de modelo de classificação binário, constantes na tabela 4.

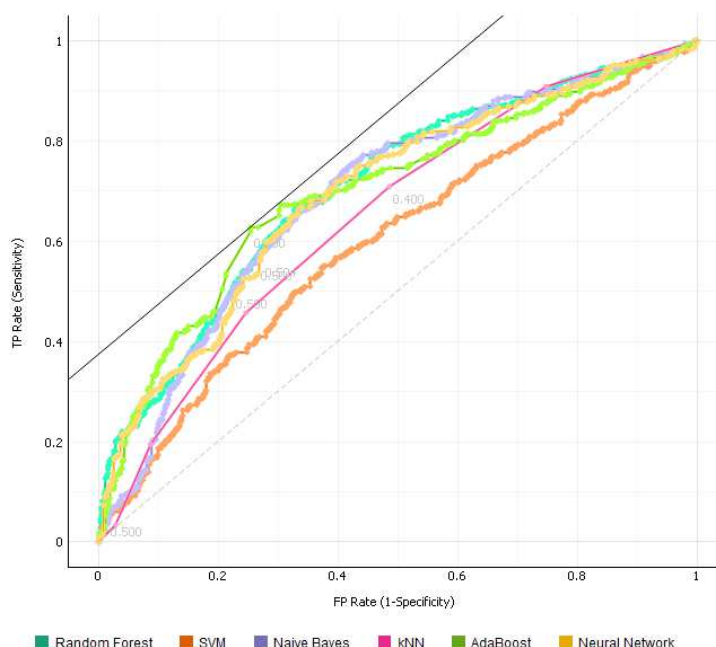


Figura 1: Curva Receiver Operating Characteristics (ROC) para comparação dos classificadores utilizados [Fonte: Orange]

Através da aplicação do filtro *AttributeSelection*, do algoritmo avaliador *CfsSubsetEval* e da estratégia de busca *BestFisrt*, chegou-se ao conjunto de parâmetros mais relevantes para o modelo de predição baseado nas variáveis listadas na Tabela 1.

Pela análise da Matriz de Confusão (*Confusion Matrix*) do modelo preditivo baseado no algoritmo *AdaBoost*, gerada com o uso do algoritmo J48 e disposta na Figura 2, observa-se que esse modelo classificou corretamente 711 (70,1%) instâncias e incorretamente 303 (29,9 %).

		Predicted		M	Σ
		Evasao	Nao Evasao		
Actual	Evasao	230	152	0	382
	Nao Evasao	151	481	0	632
M		0	1	1	2
Σ		381	634	1	1016

Figura 2: Matriz de confusão do modelo baseado no algoritmo *AdaBoost* [Fonte: Orange]

A seguir são detalhados alguns resultados e avaliações por atributos coletados, distribuídos entre as questões de pesquisa consideradas.

5.1 Atributos coletados com maior correlação com evasão em CS1

O atributo com maior correlação com evasão em CS1 para o modelo é a opção pelo curso (*A9_curso*), conforme observado durante a análise exploratória dos dados coletados. Ademais, a diferença entre os índices de evasão entre o curso de Matemática Licenciatura Matutino (61,8%) e o de Engenharia de Produção (22,2%), conforme apresentado na Tabela 3, já antecipava uma possível forte correlação.

Tal resultado parece estar vinculado à ausência de identificação ou vocação com o curso de graduação escolhido [Brito *et. al* 2015], razão pela qual a maioria dos estudantes das graduações em engenharia possam carregar consigo um possível engajamento ou vínculo mais forte com o curso escolhido.

Cabe aqui, possivelmente, uma análise futura sobre as informações de primeira e segunda opção de escolha de graduação desses estudantes no Sistema de Seleção Unificada (SISU), a fim de que seja identificado se o curso em andamento era o realmente pretendido por aqueles que foram submetidos ao referido processo seletivo.

5.2 Caracterizando a proporção entre os atributos coletados e os índices de evasão em CS1

Nos dados analisados das turmas de 2016 a 2018, 35,9% dos alunos matriculados em CS1 eram do sexo feminino e 64,1% eram do sexo masculino, com 34% de evasão observada para esse grupo de alunos. Dentre esses estudantes, 66,7% não possuíam experiência prévia em programação e 61,2% era originário do ensino médio público convencional. Esse grupo obteve um índice de evasão de 37,3%.

Do total de mulheres, 36,1% abandonou a disciplina antes do fim do período letivo, número muito próximo ao encontrado dentre os homens. Ao se combinar os mesmos atributos utilizados anteriormente para os homens, os números de evasão permaneceram próximos: 38,5%.

Percebe-se, ainda, que a combinação dos atributos **A06_origem_ensino_medio** (técnico) com o sexo feminino ou masculino e a **A03_experiencia_programacao** ou não trouxe ganhos significativos ao modelo. Por outro lado, a ausência de filhos para estudantes do sexo feminino e masculino interferiu significativamente os índices de evasão, elevando os percentuais para acima de 57% em ambos os casos.

Pôde ser observado, ainda, nesses mesmos grupos de evadidos (homens ou mulheres com filhos) que 8,6% dos alunos possuíam origem do ensino médio técnico. Do total de alunos evadidos neste conjunto, 81,5% advinha de escolas públicas convencionais e 9,9% de escolas particulares.

O atributo **A04_faixa_etaria** demonstrou inserir um excelente ganho de informação ao modelo. Para alunos entre 40 e 49 anos, o índice de evasão chega a 75,0%, enquanto que os estudantes de até 22 anos se evadiram de CS1 em apenas 32,0% dos casos, o que pode estar vinculado a responsabilidades profissionais e familiares já assumidas pelos integrantes da primeira faixa etária. Entre os cotistas e não cotistas, a diferença entre os índices de evasão foram mínimos, abaixo de 3%, descartando qualquer relação com o processo de evasão.

Visando obter uma análise mais aprofundada sobre quais atributos socioeconômicos de alunos são indicadores mais relevantes de evasão em CS1, aplicou o algoritmo baseado em árvore J48 para geração de uma árvore de decisão (*decision tree*) cujo resultado pode ser observado na figura 2, onde os nós em rosa representam a estimativa de conclusão e os nós em azul e verde a de *dropout*. A diferença nos tons de cores é devido à divisão do nó pai. Quanto mais escura a cor, maior o ganho de informação na estimativa (menor a entropia).

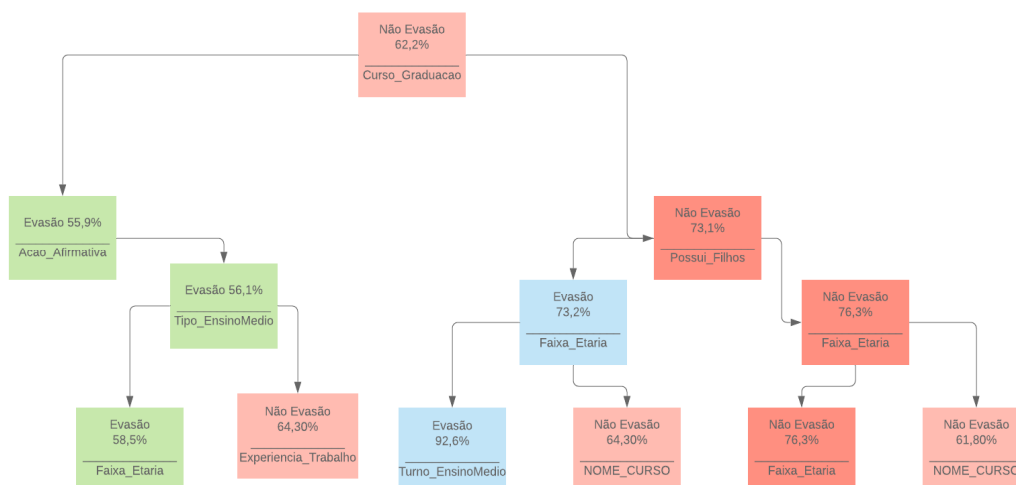


Figura 2: Arvore de decisão do modelo baseado no algoritmo *AdaBoost* [Fonte: Orange]

6. Conclusões e trabalhos futuros

O desempenho acadêmico de estudantes em *CSI* pode estar eventualmente influenciado a fatores extracurriculares. Os resultados obtidos nesta pesquisa demonstraram que, na maioria dos casos, características ligadas à dimensão social do estudante, como faixa etária e existência de filhos, influenciam tanto quanto atributos ligados à dimensão educacional como curso escolhido e origem do ensino médio.

Em suma, entender a forma pela qual este aluno interage e se faz presente na família e na sociedade, tais como responsabilidades profissionais, estrutura familiar e financeira, pode representar o ponto inicial de substituição de uma análise, não esporadicamente tangencial, encontrada em boa parte das ações direcionadas à

manutenção do quadro de alunos de disciplinas introdutórias de programação em cursos de graduação das áreas de engenharia e ciências exatas.

No decurso do trabalho, a identificação dos alunos que apresentam risco de evasão através do uso técnicas de mineração de dados mostrou-se viável. A qualidade dos resultados iniciais abre a possibilidade de novas investigações futuras, como por exemplo, o desenvolvimento de uma ferramenta de auxílio acadêmico que identifique quais alunos apresenta maior risco de abandonar os estudos de CS1.

Por esses motivos, sustenta-se a intenção de aprofundamento desta pesquisa e desenvolvimento de um modelo preditivo de evasão em CS1, com base em atributos socioeconômicos, a fim de que, entre outras questões, possam ser abordadas políticas institucionais de apoio a estudantes com risco de evasão da disciplina ou do ambiente escolar.

Como trabalhos futuros registram-se a possibilidade de coletar e analisar um número maior de atributos socioeconômicos, bem como realizar *tunning* nos classificadores e avaliar resultados de outros algoritmos de classificação e novas bases de testes objetivando a construção de um modelo probabilístico de predição de evasão antecipada em *CS1* com acurácia mais apurada.

Referências

- Alves, M. e Mantovani, K. (2016) “Identificação do perfil dos acadêmicos de engenharia como uma medida de combate à evasão” In Revista de Ensino de Engenharia. Volume 35, Número 2.
- Brito, D., Pascoal, T., Araújo, J., Lemos, M. e Rêgo, T. (2015) “Identificação de estudantes do primeiro semestre com risco de evasão através de técnicas de Data Mining.” In TISE 2015 - XX Congresso Internacional de Informática Educativa. Santiago.
- Dawson, J.Q., Allen, M., Campbell, A e Valair, A.(2018) “Designing an Introductory Programming Course to Improve Non-Majors’ Experiences.” In SIGCSE ’18: The 49th ACM Technical Symposium on Computer Science Education, Feb. 21–24, 2018, Baltimore, MD, USA. ACM, New York, NY, USA, 6 pages.
- Digiampietri, L., Nakano, F., e Lauretto, M. (2016) “Mineração de Dados para Identificação de Alunos com Alto Risco de Evasão: Um Estudo de Caso.” Revista de Graduação USP, Volume 1, Número 1.
- Fritsch, R. (2015) “A problemática da evasão em cursos de graduação em uma universidade privada”. In 37ª Reunião Nacional da ANPEd – 04 a 08 de outubro de 2015, UFSC – Florianópolis.
- Giraffa, L. e Mora, M. (2015) “Evasão na disciplina de algoritmo e programação: um estudo a partir dos fatores intervenientes na perspectiva do aluno”. Congressos CLABES III. México.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. e Witten, I. (2009) “The weka data mining software: an update”. In ACM SIGKDD explorations newsletter, ACM.
- Júnior, J., Noronha, R. e Kaestner, C. (2017) “Método de Seleção de Atributos Aplicados na Previsão da Evasão de Cursos de Graduação”. In Revista de Informática Aplicada, Volume 13, Número 2.

- Pappas, I., Giannakos, M. and Jaccheri, M. L. (2016). “Investigating Factors Influencing Students' Intention to Dropout Computer Science Studies.” ITiCSE '16. Arequipa, Peru: ACM.
- Pereira, F., Oliveira, E., Cristea, A., Fernandes, D., Silva, L., Aguiar, G., Alamri, A. e Alshehri, M.(2019) “Early Dropout Prediction for Programming Courses Supported by Online Judges.” In Isotani S., Millán E., Ogan A., Hastings P., McLaren B., Luckin R. (eds) Artificial Intelligence in Education. AIED 2019. Lecture Notes in Computer Science, vol 11626. Springer, Cham
- Petersen, A., Craig, M., Campbell, J. e Tafliovich, A. (2016) “Revisiting why students drop CS1”. In Proceedings of the 16th Koli Calling International Conference on Computing Education Research (Koli Calling '16). ACM, New York, NY, USA, 71-80.
- Quille, k., Culligan, N. e Bergin, S. (2017) “Insights on Gender Differences in CS1: A Multiinstitutional, Multi-variate Study.” ITiCSE '17, July 03-05, 2017, Bologna, Italy. ACM
- Ramos, J., Silva, J., Prado, C., Gomes, A. e Rodrigues, R. (2018) “Um estudo comparativo de classificadores na previsão da evasão de alunos em EAD”. In VII Congresso Brasileiro de Informática na Educação (CBIE 2018) Anais do XXIX Simpósio Brasileiro de Informática na Educação (SBIE 2018)
- Rêgo, D., Pascoal, T. e Araújo, e J. (2015) “Identificação de estudantes do primeiro semestre com risco de evasão através de técnicas de Data Mining.” In TISE 2015 - XX Congresso Internacional de Informática Educativa. Santiago.
- Santana, B. L., Figuerêdo, J. S. L. e Bittencourt, R. A. (2017) “Motivação de Estudantes Non-Majors em uma Disciplina de Programação”. In WEI 2017 - XXV Workshop sobre Educação em Computação. São Paulo.
- Wasik, S., Antczak, M., Laskowski, A. e Sternal, T. (2018).“A Survey on Online Judge Systems and Their Applications”. In ACM Computing Surveys. New York
- Witten, I., Frank, E. e Hall, M. (2011) “Data Mining: Practical Machine Learning Tools and Techniques.” 3rd. ed.: Morgan Kaufmann Publishers Inc. San Francisco CA, USA.