

IA: DESCRIÇÃO E APLICAÇÃO DE REGRAS DE EVASÃO NO CURSO DE CIÊNCIA DA COMPUTAÇÃO EM IES¹

IA: DESCRIPTION AND APPLICATION OF EVASION RULES IN HEI COMPUTER SCIENCE COURSE

Crhistopher Lenhard² e Mirkos Ortiz Martins³

RESUMO

Esta pesquisa tem como objetivo definir regras para rastrear, em uma base de dados acadêmica, padrões de comportamento e características de alunos evasores em curso de Ciência da Computação de Instituição de Ensino Superior (IES). Foram utilizados dados gerais de um período de 2008 a 2018 contendo 1173 alunos matriculados e analisados 427 alunos evasores, armazenados em banco de dados MS SQL Server. Foram levantadas sete regras que classificam de forma razoável os desistentes, a partir de consultas SQL, e alimentam uma lógica de predição por inteligência artificial através de programação em linguagem Python. Os alunos evasores geraram perfis dentro das sete regras, que tornam possível encontrar entre discentes quais são os possíveis desistentes de curso e assim iniciar ações contra esse tipo de ocorrência na vida acadêmica do aluno.

Palavras-chave: aprendizado de máquina, busca perfil, Python.

ABSTRACT

This research aims to define rules for tracking, in an academic database, behavioral patterns and characteristics of students evading undergraduate computer science from Higher Education Institution (HEI). General data were used over a period of 10 years - 2008 to 2018 - containing 1674 students enrolled and analyzed 427 dropout students, stored in MS SQL Server database. The defined rules reasonably classify the quitters, from SQL queries and Python programming language. The student dropouts generated profiles within seven rules, which made it possible to find among students which are those possible dropouts of the course and thus initiating actions against this type of occurrence in a student's academic life.

Keywords: machine learning, profile search, python.

¹ Trabalho de Iniciação Científica - Projeto de Pesquisa CNPQ, edital 30/2017.

² Acadêmico do curso de Ciência da Computação - UFN. Bolsista CNPQ. E-mail: crhistopher.lenhard@ufn.edu.br

³ Orientador. Docente do curso de Ciência da Computação - UFN. E-mail: mirkos@gmail.com

INTRODUÇÃO

O desenvolvimento de um sistema de preempção (previsão) de evasão de discente em uma Instituição de Ensino Superior (IES) é de extrema importância para a manutenção da estrutura funcional diretamente ligada aos cursos oferecidos, tais como docentes, secretariado e coordenação, pois permite a definição de estratégias de manutenção do número de alunos após a matrícula e a diminuição do fechamento de turmas ou mesmo turmas com baixo número de alunos.

Muitas vezes a gerência e coordenação de um curso do ensino superior envolve muitas variáveis e dentre essas variáveis, aquelas que determinam um provável aluno evasor podem passar despercebidas pela administração acadêmica - seja pelo grande volume de dados envolvidos na coordenação, seja pelo desconhecimento de quais são as características importantes de serem observadas para fazer uma previsão do tipo.

Nesse contexto, esse trabalho pretende levantar uma série de regras de identificação de dados para serem utilizados como entrada em uma ferramenta de *data science*, estado da arte na computação científica, para a descoberta de padrões nos diversos dados em uma base de informações acadêmicas e apoiar em decisões estratégicas para a manutenção de alunos prováveis evasores do sistema de ensino da instituição.

A evasão escolar é um problema enfrentado em várias Instituições de Ensino Superior (IES), sejam elas públicas ou particulares. Enquanto vários esforços são empregados para diminuir a desistência de alunos, esses são divididos em duas frentes: encontrar padrões nos perfis dos alunos que abandonam seus cursos e desenvolvimento de campanhas para fidelizar alunos com probabilidade de evadir da IES (TINTO, 1999).

Na avaliação da provável desistência de um aluno, é preciso considerar e encarar de forma mais individualizada o sucesso acadêmico, medido na forma de andamento do curso, objetivando atingir métricas mais concretas na interação IES - aluno e assim definir um formato mais dinâmico para o acompanhamento de sua vida acadêmica possibilitando a preempção de provável desistência do respectivo curso (FILHO *et al.*, 2007). Também é um complicador desse contexto de desistência, o perfil de cada aluno confrontado com as características particulares de cada curso, onde o casamento entre os diferentes atores nem sempre é possível de um sucesso (DEKKER; PECHENIZKIY; VLEESHOUWERS, 2009).

No contexto de solução desse problema, surge o desenvolvimento de uma aplicação utilizando *machine learning*, ou aprendizado de máquina (GRUSS, 2016), para auxiliar o gestor na identificação das características de um aluno provável evasor, antes que ele realmente desista do curso, para que através de estratégias desenvolvidas pela área administrativa do curso e da Instituição, haja uma tentativa de contorno do cenário de desistência. Um passo importante para a implantação de um sistema indicador de evasão é a identificação das regras que regem o comportamento de desistência do aluno, analisados sob a ótica de dados armazenados em informações acadêmicas.

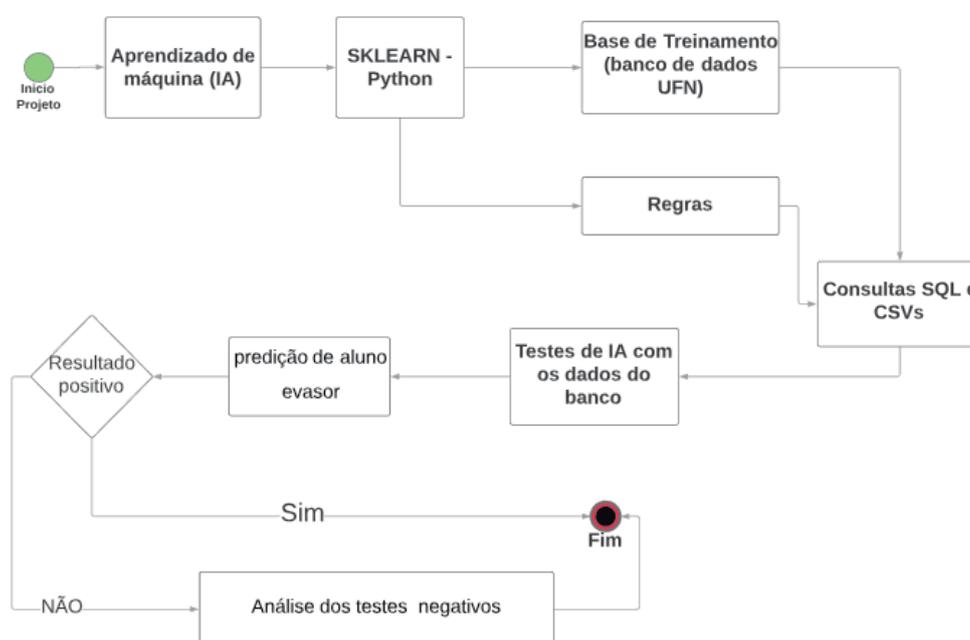
Os objetivos específicos desse trabalho se aplicam em estruturar um conjunto de regras para modelar uma série de algoritmos de inteligência artificial (IA) (RUSSEL; NORVIG, 2013), para a verificação daquele código que melhor se aplica ao problema de evasão discente, assim como modelar a classificação dos dados do problema de evasão no curso de ciência da computação, fazer a elaboração de regras utilizadas para reconhecimento de alunos evasores e posteriormente ampliar a solução para diversos cursos do ensino superior e modelar a base de dados da UFN como fonte de informações, para o sistema desenvolvido e analisar os resultados de evasão institucional. Também se objetiva a capacitação de discentes (alunos envolvidos com o desenvolvimento do algoritmo, sejam em iniciação científica ou mesmo trabalhos finais de graduação) para a construção de soluções usando *data science*.

MATERIAL E MÉTODOS

A metodologia de desenvolvimento do projeto pode ser visualizada na Figura 1, onde se identifica a fase inicial de definição de métricas de aprendizado de máquina (*machine learning*) para a listagem de quais as possíveis implementações podem ser utilizadas para a solução do problema proposto. Em seguida, é escolhida a tecnologia para resolver a programação, baseada em curva de aprendizado, eficiência, desempenho e documentação. Optou-se por construir a solução de busca de evasores utilizando a linguagem de programação Python em conjunto com a biblioteca *Sklearn*.

O passo seguinte é a compreensão do mapa de dados da base acadêmica de informações da IES, utilizando-a como treinamento para os algoritmos de inteligência artificial. A base de dados acadêmica obtida lista alunos do curso de ciência da computação no período de 2008 a 2018, totalizando 1173 alunos analisados.

Figura 1 - Metodologia do trabalho.



Ainda na figura 1, em paralelo ao estudo do mapa de dados da base acadêmica, a definição das regras - objeto desse trabalho - ocorre para o ajuste fino e classificação de alunos entre evasores e não-evasores, servindo de conhecimento para algoritmos de inteligência artificial e assim reconhecer perfis de alunos no que tange a permanência no curso de graduação.

RESULTADOS E DISCUSSÃO

Considerando a fase de definição das regras para classificação dos alunos (em prováveis evasores, evasores e não evasores), foram identificadas diferentes regras, descritas na seção seguinte. Todas as regras levaram em consideração o mapa de dados da base acadêmica e representam uma aplicação de lógica de localização de evasores, consistente com (FILHO *et al.*, 2007).

IDENTIFICAÇÃO DE REGRAS

Antes da posse dos dados acadêmicos, estudou-se conceitualmente regras gerais para identificar evasores e após a análise da base de dados acadêmica fornecida, foram identificadas mais algumas regras específicas para identificação de prováveis evasores (*pe*), também listadas na tabela 1:

- Idade
- Ambiente (Gênero)
- Evasor anterior
- Curso como segunda opção
- Reprovação versus semestre
- Média individual versus média turma
- Problema Financeiro

Tabela 1 - Regras de normalização dos alunos.

Regra	Descrição	Equação	Decisão
Problema Financeiro	Identificar um aluno que deixa de pagar mensalidades. A função deve identificar nos evasores quantos meses eles deixaram de pagar.	Estavam em dia? Quantos meses eles não pagaram?	A partir do mês N-1, identificar possível evasor.
	Identificar um aluno que deixa de pagar mensalidades. A função deve identificar nos evasores quantos meses em sequência eles deixaram de pagar.	Quantos meses eles não pagaram em sequência?	A partir de quantos meses é possível identificar um evasor.
Reprovação versus Semestre	Identificar nos evasores aqueles que tiveram um número de reprovações nos primeiros semestres	Identificar as reprovações (cumulativas), cria um índice vezes o semestre (menor semestre tem maior peso)	Encontrar o semestre e o número de reprovações onde é disparada a evasão.

Média individual	Média individual for muito abaixo da média da turma	Média da turma comparada com a média individual	Identificar perfil do aluno em relação às médias em diferentes disciplinas.
Curso como segunda opção	Identificar se o curso foi uma segunda opção para o aluno	Atribuir peso maior para um aluno que tem como selecionado seu curso como segunda opção para a decisão de um <i>pe</i>	Idem a equação
Aluno evasor anteriormente	Identificar se o aluno tem em seu histórico escolar passagens por outros cursos.	Atribuir peso maior para um aluno com ficha de evasor anterior para a decisão de um <i>pe</i>	Idem a equação
Idade	A idade de um aluno influencia na sua decisão para um <i>pe</i>	Idade do aluno e seu presente semestre comparados a um padrão de evasores	Identificar um <i>pe</i> pela sua idade e semestre
Ambiente (gênero feminino)	Uma mulher tem uma maior afinidade de evasão em ciência da computação	Atribuir peso maior para um aluno do gênero feminino	Idem a equação

A regra de idade (f) analisa o quanto influencia essa característica na decisão de evasão para um aluno *pe*. Usa-se a idade do aluno e seu semestre atual comparado com as idades dos alunos evasores já identificados na base acadêmica:

$$(idade \wedge semestre) \in padraoEvasor \Rightarrow peso = peso_evasor \vee peso = 0 \tag{1}$$

onde o peso igualado a zero indica um aluno que não é *pe*.

A regra (g), que leva em consideração o gênero do aluno no momento da matrícula - possíveis masculino e feminino nessa base de dados - atribui pesos para um aluno *pe* conforme a probabilidade levantada em análise dos evasores já identificados:

$$genero == feminino \Rightarrow peso = 1 \boxtimes peso = 0 \tag{2}$$

onde o gênero feminino aumenta consideravelmente a probabilidade de um aluno *pe* de tornar-se efetivamente um evasor.

Na regra que identifica se o aluno é evasor anterior (e) de algum outro curso, identifica-se se uma determinada pessoa (o aluno) possui outro registro de matrícula na base de dados acadêmica, conforme:

$$aluno_historico == possui_evasao \Rightarrow peso = 1 \vee peso = 0 \tag{3}$$

que se faz a análise direta se já houve evasão ou não. Uma evolução dessa regra poderia atribuir peso conforme o número de evasões anteriores, mas em estudo da base acadêmica alvo não foram encontrados alunos *pe* com mais de uma evasão anterior, logo a regra não precisou avaliar um número maior que esse de evasões.

Para a avaliação de um aluno motivado em cursar o seu atual curso, identifica-se se o aluno optou originalmente por ciência da computação ou esse curso foi sua segunda opção (regra d) caracterizando a seguinte proposição:

$$escolha_curso == segunda_opcao \Rightarrow peso = 1 \vee peso = 0 \quad (4)$$

com classificação semelhante à regra c, onde se ignora mais ocorrências de segunda opção para um mesmo aluno *pe* dentro da base de dados acadêmica.

Na regra (b), avalia-se o aluno conforme seu aproveitamento acadêmico baseado em suas reprovações. A regra “Reprovação versus Semestre”, responsável em identificar número de reprovações (que podem ser sequenciais ou não) em um semestre e ao decorrer do curso pode ser descrita por:

$$\alpha = \sum_{i=1}^8 P_i * disciplina_reprovada_i \quad (5)$$

com $P_i = 9 - i$, a equação baseia-se em somar o número de reprovações com o custo do peso de reprovação de cada semestre e multiplicar disciplina reprovada (que possuem diferentes pesos pelo andamento do curso). Podem ser identificadas quais as disciplinas que mais reprovam e em quais semestres alunos evasores tendem a evadir por conta do número de reprovações. Disciplinas de semestres iniciais possuem pesos maiores e disciplinas de fim de curso possuem pesos menores na determinação do índice α .

A regra (c) que compara a média individual do aluno com a média de sua turma, procura encontrar características de nivelamento de turma e prováveis desvios que possam influenciar o aluno na decisão de evadir do curso. A média da turma calculada por:

$$Media_{turma} = \sum_{i=1}^{nAlunos} \frac{media_{aluno_i}}{nAlunos} \quad (6)$$

determinada anteriormente a nota de todos os colegas do aluno o normaliza por:

$$Media_{aluno} \leq (Media_{turma} * 0,75) \Rightarrow pe \vee npe \quad (7)$$

para classificá-lo conforme sua nota em relação à turma. Se a média individual for menor que 75% da média da turma, então o aluno é um *pe*, caso contrário ele é um aluno não provável evasor *npe*.

Por último, a regra (a) que relaciona problemas financeiros ao aluno *pe* analisa as mensalidades atrasadas (*m_atr*) que são limitadas a seis por semestre, segundo o regulamento da instituição, conforme a equação:

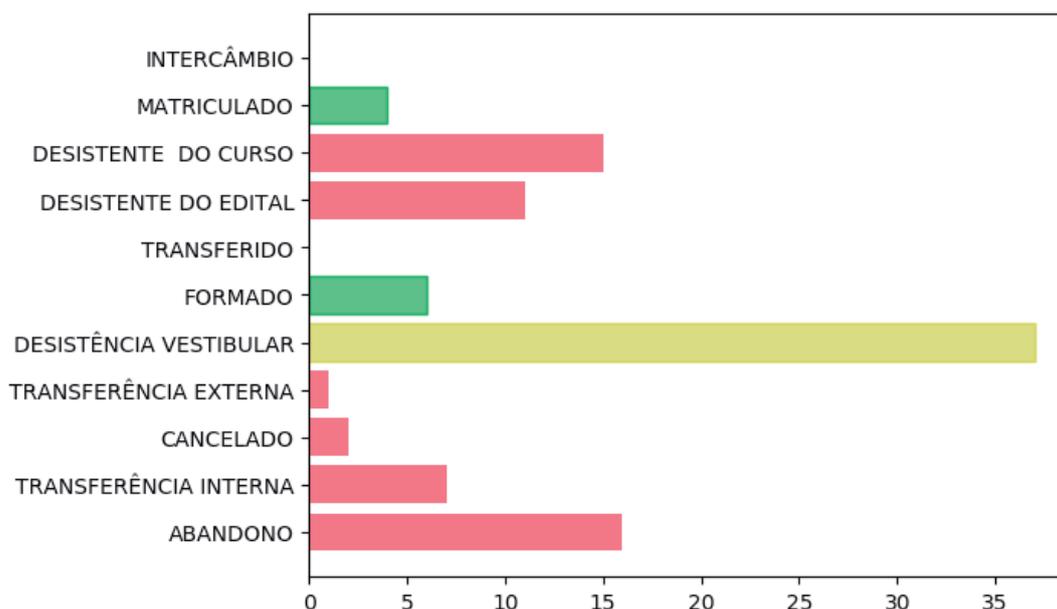
$$\alpha = \sum_i^6 m_atr * i \quad (8)$$

Ao final do levantamento das regras, analisando-se a base de dados institucional e aplicando as regras foram obtidos os seguintes resultados:

- a) 1173 alunos cadastrados para o curso de Ciência da Computação no período compreendido entre 2008 e 2018 na Instituição;
- b) 1074 alunos do sexo masculino e 99 alunas cadastrados;
- c) 376 alunos, entre homens e mulheres, já desistiram após o vestibular e nem iniciaram a cursar as disciplinas, não contabilizando como evasão propriamente dita;
- d) 198 alunos estão em situação matriculados, formados ou intercâmbio

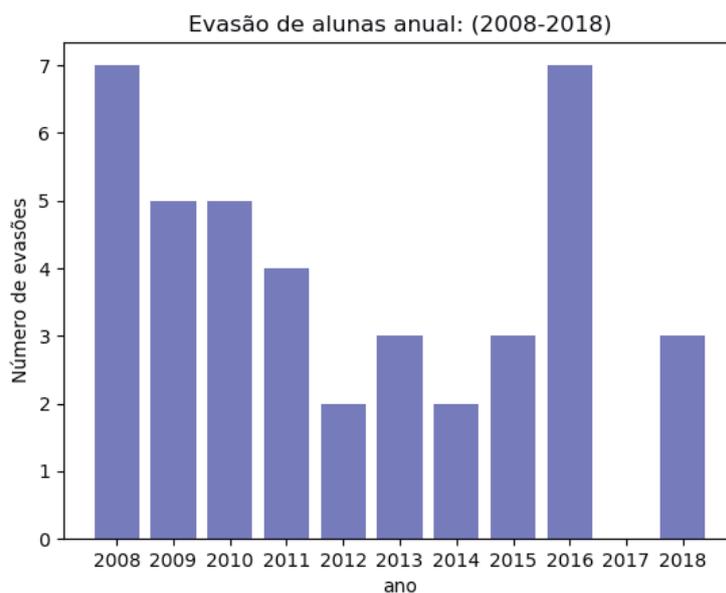
Na figura 2 são mostrados os *status* do conjunto de 99 mulheres cadastradas na base de dados da Instituição para o período de 2008 a 2018 referentes ao curso de Ciência da Computação. Em amarelo, listam-se as alunas que desistiram já no vestibular (37 no total) que nem foram contabilizadas, nesse trabalho, como evasão discente visto que elas nem chegaram a cursar uma única disciplina no curso. Na mesma figura, em verde, estão os *status* onde não temos representada uma evasão, que são os casos de intercâmbio (0 aluna), matriculado (4 alunas) e formadas (6 alunas).

Figura 2 - Status das mulheres (2008-2018).

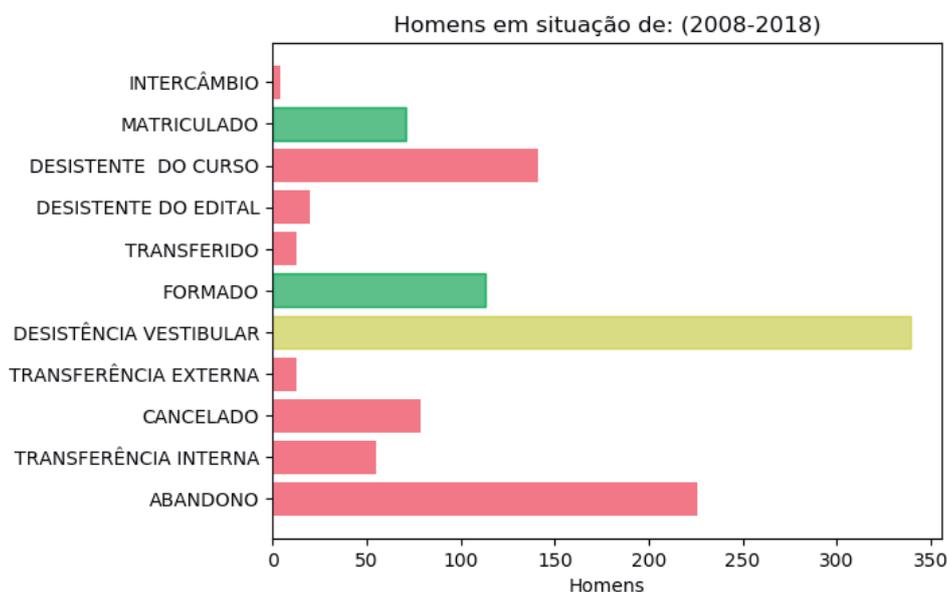


Na cor vermelha estão todas aquelas situações onde as alunas não concluíram o curso: abandono (16 alunas), transferência interna (7 alunas) e externa (1 aluna), cancelamento (2 alunas), desistência de edital (11 alunas) e de curso (16 alunas). Nenhuma aluna foi transferida ou entrou em intercâmbio nesse período de estudo, no curso de Ciência da Computação na Instituição.

Na figura 3, é possível observar uma variação entre 2 e 7 desistências por ano, exceto no ano de 2017, entre as alunas do curso de computação.

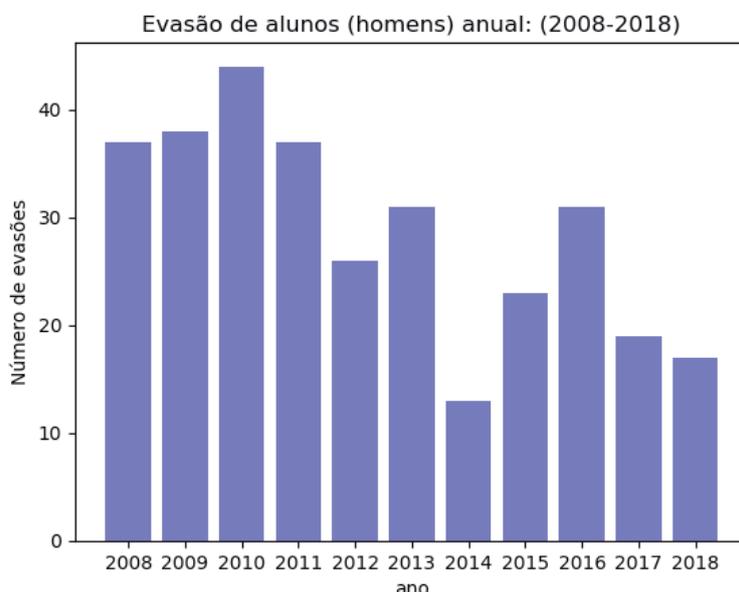
Figura 3 - Evasões ano a ano no período.

Em relação aos alunos, é mostrado o gráfico representativo na figura 4, onde são listados da mesma forma todas as formas encontradas de representação de status para os alunos homens no período. São mostrados os 339 alunos que desistiram após o vestibular, não contabilizados como evasores, pois não chegaram a cursar nenhuma disciplina do curso de Ciência da Computação. Em verde são apresentados aqueles casos que não contabilizam evasão: formados (113 alunos), matriculados (71 alunos) e em intercâmbio (4 alunos). Na cor vermelha são apresentados aqueles casos onde os alunos são evasores: abandono (226 alunos), transferência interna (55 alunos) e externa (13 alunos), cancelaram a matrícula (79 alunos), transferidos (13 alunos) e aqueles que desistiram via edital (20 alunos) ou do curso (141 alunos).

Figura 4 - Homens conforme dados de status na Instituição.

Um detalhe interessante nos dados analisados é a informação que todos os casos de intercâmbio ocorreram com alunos do sexo masculino, assim como o *status* na base de dados do tipo transferência, talvez referenciando a uma transferência militar, como podemos observar na figura 5.

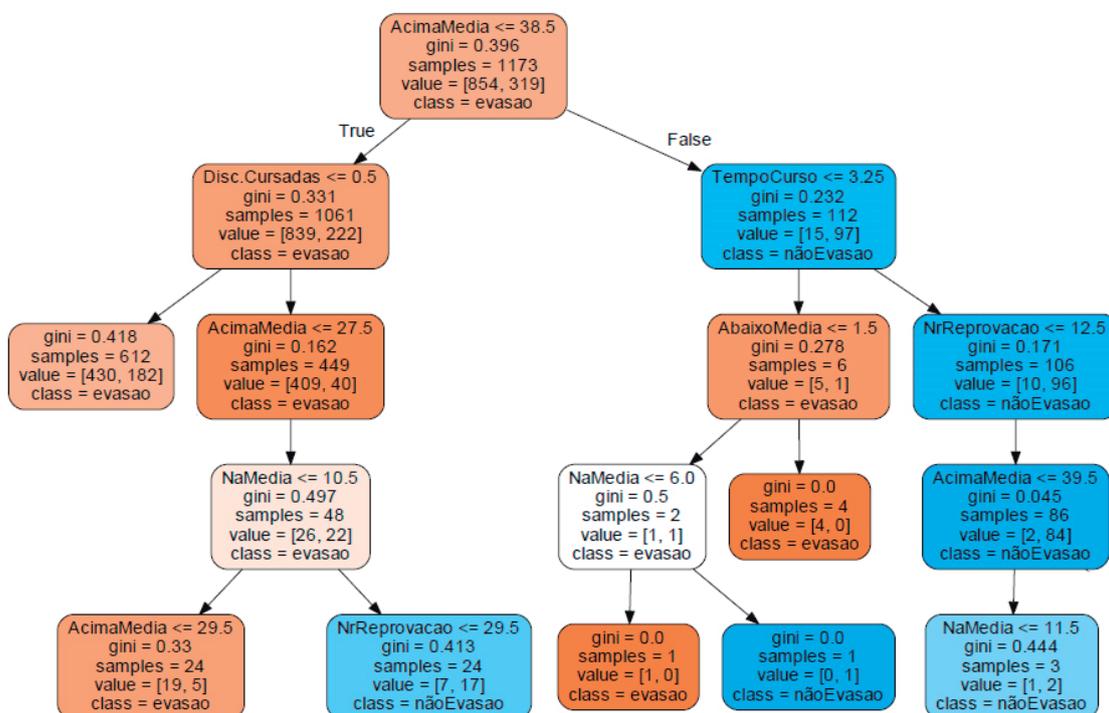
Figura 5 - Evasão ano a ano de homens (2008-2018).



Conforme pode ser observado na figura 5, o número de evasores homens no curso de Ciência da Computação diminuiu, também por questão do menor número de inscritos, porém percentualmente a proporção também diminuiu.

A árvore de decisão gerada pela implementação com Python e *sklearn* (árvore de decisão) e testada com os dados fornecidos pode ser visualizada na figura 6.

Figura 6 - Trecho da Árvore de decisão gerada pelo classificador sklearn.



Neste trecho do classificador pode-se observar o funcionamento para tomada de decisão do classificador, cada etapa se ramifica em dois e as folhas da árvore representam sua classificação, onde as folhas demarcadas azuis são *npe* e as laranjas são *pe*.

CONCLUSÃO

Após a construção do algoritmo de árvore de decisão e seu uso, com a base de dados acadêmica da Universidade Franciscana, restritas ao curso de Ciência da Computação, no período de 2008 a 2018, os resultados permitiram a obtenção de perfis bem característicos dos alunos evasores. A partir dessa informação foi possível identificar elementos fortes para a predição de um provável evasor do curso. Assim, baseando-se nas sete regras listadas durante o trabalho, conclui-se que um aluno será evasor se:

- a) O número de média das notas de disciplinas abaixo da média da turma é superior a 25% das notas;
- b) O número de reprovações é maior que a 12.5 % das disciplinas cursadas pelo aluno;
- c) O tempo de curso para um aluno com notas abaixo da média da turma for superior a 3 semestres cursados;
- d) Alunas com médias abaixo da turma.

Neste contexto, nota-se que existem indicadores para o alerta de um aluno quase desistente e assim permite-se a criação de métricas para seu acompanhamento. No contexto de notas dos alunos, torna-se importante verificar a discrepância entre as notas individuais *versus* as notas médias de toda uma turma, formando um indicador de deslocamento do aluno ou mesmo desmotivação em relação aos colegas. No quesito médias também fica evidente que os alunos com melhores notas dificilmente serão evasores, salvo alguns casos de transferência interna/externa, o que indica um desvio de vocação na escolha do curso.

Para o tempo de permanência no curso foi verificado que aqueles alunos evasores em semestres iniciais - pouco tempo de curso - são identificados por uma massiva reprovação em disciplinas, ditas duras: aquelas que envolvem matemática e conceitos abstratos, sem ligação direta com a computação nos semestres iniciais. Alunos com evasão após uma permanência maior no curso também possuem um número alto de repetência (maior que 50% das disciplinas cursadas) porém com a atenuação da média geral em função da soma de algumas aprovações.

No quesito gênero, foi verificada a maior evasão feminina - comparativamente - com altos índices de médias individuais abaixo da média da turma, talvez motivada pela falta de empatia com o público masculino, falta de suporte (nenhuma das mulheres apresentou desligamento para fazer intercâmbio) e ausência de um grupo feminino para adensamento. O problema da desistência feminina na área de computação é conhecido e os números somente colaboram na construção de um cenário que deve ser atacado para aproximar as mulheres da área tecnológica e fomentar sua presença cada

vez maior, diminuindo a desistência e mostrando a capacidade em tornarem-se bacharéis em Ciência da Computação.

Das regras citadas no presente trabalho, a que não foi possível sua total aplicação foi a regra financeira, visto o embaraço na obtenção e manipulação dos dados institucionais. Esses serão alvo de criteriosa análise e aplicação em trabalho futuro (muito próximo) para o fechamento do entendimento de evasão e construção de soluções para evitar a perda de material humano nas salas de aula do curso de Ciência da Computação, bem como nos demais cursos da Universidade Franciscana.

REFERÊNCIAS

DEKKER, G.; PECHENIZKIY, M.; VLEESHOUWERS, J. Predicting students drop out: a case study. **Computers, Environment and Urban Systems**, v. 33, p. 41-50, 2009.

FILHO, S. *et al.* A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**, v. 37, n. 132, p. 641-659, 2007.

GRUSS, J. **Data Science do zero**. Rio de Janeiro: Alta Books, 2016.

RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. Rio de Janeiro: Elsevier, 2013.

TINTO, V. Taking retention seriously: Rethinking the first year of college. **NACADA Journal**: Fall, v. 19, n. 2, p. 5-9, 1999.

