

# Detecção precoce de evasão em cursos de graduação presencial em Computação: um estudo preliminar

Leandro S. G. Carvalho, Ana L. M. Santos, Fabíola G. Nakamura, Elaine H. T. Oliveira

Instituto de Computação (IComp) – Universidade Federal do Amazonas (UFAM) –  
Manaus – AM – Brasil

{galvao, analuciamachado, fabiola, elaine}@icompu.ufam.edu.br

**Resumo.** *Esta pesquisa inicial teve o objetivo de identificar atributos de estudantes, detectáveis nas duas primeiras semanas do curso, que apresentam forte correlação com parâmetros levantados na literatura como associados a risco de evasão de curso. Trata-se de uma pesquisa exploratória, de abordagem quantitativa, caracterizada como pesquisa de avaliação, delimitada a três cursos de graduação presenciais, na área de Computação, com ingressantes no período 2012-2018, na Universidade Federal do Amazonas (UFAM). Observou-se que os seguintes atributos estão significativamente associados a um bom coeficiente de rendimento no primeiro período letivo: participação das atividades de acolhimento dos calouros, política afirmativa, modalidade de Ensino Médio, conhecimento prévio de linguagem de programação, e escolha do curso como primeira opção.*

**Abstract.** *This initial work aimed to identify attributes of students, detectable in the first two weeks of the undergraduate program, that present a strong correlation with parameters identified in literature as associated with risk of evasion. This is an exploratory research, quantitative approach, characterized as evaluation research, delimited to three undergraduate programs in the area of Computing, with incoming students from 2012 to 2018, at Federal University of Amazonas (UFAM). It was observed that the following attributes are significantly associated to a good GPA in the first term: participation of welcome activities for freshmen, affirmative policy, modality of High School, previous knowledge in programming, and choice of the program as first option.*

## 1. Introdução

As Universidades exercem o papel de formar profissionais de nível superior, contribuindo para a construção de uma sociedade mais justa, disseminando e produzindo conhecimento e desenvolvimento tecnológico, econômico e social (Matos *et al.* 2019). Quanto mais alunos finalizam seus cursos e se tornam profissionais qualificados e preparados para o mercado de trabalho, maior será a contribuição social das Universidades. Por isso, a preocupação com a evasão nas Universidades Federais de Ensino, foco desta pesquisa, é objeto de muitos trabalhos de pesquisa que buscam identificar e encontrar soluções para essa questão que é recorrente e não se esgota.

Neste estudo, o objetivo é determinar quais indicadores acadêmicos, sociais, econômicos e demográficos dos calouros dos cursos da área de computação estão associados à situação final no curso: formado ou desistente. É importante observar que o estado de “desistente” representa um caso de insucesso do ponto de vista da coordenação do curso, e não do estudante. Alguns destes podem desistir de um curso pelo qual descobrem não ter afinidade e apresentar um excelente desempenho em outro com o qual se identifique.

Este estudo foi orientado pelas seguintes questões de pesquisa:

**QP1:** a partir do histórico acadêmico dos alunos de computação, quais indicadores, passíveis de serem coletados no 1º período letivo, possuem associação com o estado final (formado ou desistente) do estudante nos cursos de Computação?

**QP2:** que itens do questionário sócio-econômico-demográfico possuem maior associação com os indicadores encontrados na QP1?

## **2. Trabalhos Relacionados**

Nas análises do estudo realizado com os alunos evadidos dos cursos de Computação da Universidade de Brasília (UnB), no período de 2005 a 2015, Azevedo *et al.* (2015) concluíram que a escolha do curso e a falta de base em conceitos do ensino médio estão dentre os fatores causadores da evasão.

Para Azevedo *et al.* (2015) e Hoed (2016), muitos alunos ingressam no curso sem terem a real concepção do que é ser um profissional da área de computação. Somente o fato de gostarem de computador gera nos ingressantes a falsa convicção de que têm vocação para a área, ilusão que logo se desfaz ao se depararem com as disciplinas específicas de cálculo e de programação, associado ao fato de alguns não possuírem a base necessária requerida para o entendimento dessas disciplinas. Além desses fatores, a pesquisa também detectou que os picos de evasão ocorrem nos períodos iniciais dos cursos de Computação. De modo semelhante, Morães e Pombeiro (2016) encontraram que um dos motivos da evasão é o não conhecimento prévio dos objetivos do curso escolhido.

Além de informações prévias obtidas de questionários e dados estáticos obtidos dos alunos, Manhães *et al.* (2011) e Rigo *et al.* (2012) apontam a importância de um amplo mapeamento de fatores associados à evasão, por meio de técnicas de mineração de dados, para a implantação de soluções interativas que possibilitem o diagnóstico precoce e a realização de ações pedagógicas relevantes.

Nesse sentido, o presente estudo busca identificar indicadores, possíveis de serem coletados nas duas primeiras semanas de aula do curso, que possuam forte associação com o desempenho dos estudantes no 1º período letivo do curso e, consequentemente, com o desempenho final do aluno, ou seja, se este se forma ou se desiste do curso em que ingressou.

### **3. Metodologia**

Nesta seção, descrevemos o instrumento elaborado para levantar as características socioeconômicas e demográficas dos ingressantes dos cursos de Computação a partir de 2018. Em seguida, são descritas as etapas de realização do presente estudo.

#### **3.1 O instrumento**

O questionário de caracterização sócio-econômico-demográfico (CSED) utilizado nesta pesquisa foi elaborado a partir de questionários semelhantes aplicados aos alunos que participam do Enade, Enem e processos seletivos de outras universidades. As perguntas foram selecionadas pensando em obter o máximo de informação sobre a situação sócio-econômica e demográfica dos ingressantes nos cursos de computação. A redação das perguntas e das opções de respostas foi baseada nas diretrizes de Günther (2003).

O objetivo principal do questionário é levantar mais informações sobre os alunos dos cursos de computação, de forma censitária, a fim de permitir às Coordenações de Curso se anteciparem a uma possível situação de evasão e proporem, dentre as alternativas oferecidas pela Instituição de Ensino, uma saída para o problema apresentado. Lidar com uma situação cujo diagnóstico é previamente conhecido permite buscar solução, sem precisar ser invasivo ao perguntar diretamente ao aluno sobre sua situação socioeconômica e demográfica.

O questionário<sup>1</sup> é composto por 38 (trinta e oito) questões de múltipla escolha e é disponibilizado em um AVA da instituição, baseado no Moodle. É aplicado nas duas primeiras semanas do 1º período letivo, com a presença dos pesquisadores, em data, horário e laboratório previamente acordado com os professores das disciplinas que cedem um tempo de suas aulas para que os alunos possam preencher o questionário. Este é iniciado apenas após os estudantes concordarem com o Termo de Consentimento Livre e Esclarecido, também disponível no AVA. O preenchimento é voluntário, mas nos dois anos de aplicação (2018 e 2019), nenhum aluno presente fez objeção ao preenchimento do questionário. Na realidade, alguns até contribuíram apresentando sugestões para aprimoramento do instrumento.

A pesquisa foi submetida à apreciação do comitê de ética em pesquisa com seres humanos, que atestou a não aplicabilidade de autorização por se tratar de estudo de natureza censitária.

#### **3.2 Procedimentos**

Uma vez que o questionário CSED foi aplicado pela primeira vez em 2018, não havia dados disponíveis para investigar a relação entre os indicadores do questionário, coletados nas duas primeiras semanas do curso, e a situação final, que somente será conhecida dentro de quatro anos, pelo menos, pois é a duração prevista para dois dos três cursos acompanhados. Dessa forma, para contornar essa impossibilidade, o estudo foi dividido em duas etapas:

---

<sup>1</sup> Versão que dispensa cadastro no AVA disponível em <https://github.com/leandrogalvao2/wei2019>

1. Classificação da situação final do aluno (formado ou desistente) com base em atributos coletáveis após o final do 1º período letivo.
2. Análise da associação entre os indicadores do questionário CSED e os parâmetros identificados na etapa 1 como melhores estimadores da situação final do aluno.

Os procedimentos metodológicos utilizados em cada etapa são descritos nas subseções a seguir.

### 3.2.1 Classificação da situação final do aluno com base no histórico

Nesta etapa do estudo, foi extraído o histórico de alunos que correspondiam às seguintes condições:

1. Ter registro de matrícula nos cursos de Ciência da Computação, Engenharia de Software e Engenharia da Computação, os únicos da área de Computação oferecidos pelo campus Manaus da Universidade Federal do Amazonas UFAM).
2. Ter ingressado no curso nos anos de 2012 a 2017.
3. Ter situação final igual a uma das seguintes categorias: “formado”, “desistente” ou “jubilado”. Na fase de tratamento de dados, essas duas últimas classes foram tratadas como sendo de uma única classe: “desistente”.

O ano de 2012 marcou o início da adoção de cotas nos processos seletivos da UFAM, em conformidade com a Lei nº 12.711/2012. Por isso, ele foi adotado como ano de início da coleta dos dados. No estado do Amazonas, a maior parte da população se auto-declara preto, pardo ou indígena. Consequentemente, há poucos estudantes que ingressam nas vagas de cor não declarada. Assim, neste estudo, foram consideradas apenas quatro categorias de cota: **ampla concorrência** (AC); **cota dependente de renda** (renda familiar bruta per capita igual ou inferior a 1,5 salário mínimo que tenha cursado integralmente o ensino médio em escolas públicas, tendo ou não declarado cor da pele); **cota independente de renda** (sem declaração de renda e que tenha cursado integralmente o ensino médio em escolas públicas, tendo ou não declarado cor da pele); e **outros** (ingresso via mandado judicial, transferência ex-offício ou vagas remanescentes).

Por fim, os seguintes atributos foram coletados para cada aluno, a fim de alimentar os algoritmos de classificação:

- **CR-1P** (coeficiente de rendimento do 1º período): média aritmética das notas das disciplinas cursadas durante o primeiro período letivo.
- **TX-AP-1P** (taxa de aprovação em disciplinas de 1º período): razão entre o número de disciplinas aprovadas e o total de disciplinas cursadas pelo aluno durante o primeiro período letivo.
- **COTA**: valor pertencente a uma das quatro categorias de cota consideradas.
- **NotaIP**: nota final em disciplina de Introdução à Programação (IP).
- **NotaAL1**: nota final na disciplina Álgebra Linear I (AL1).

Não foram utilizadas as notas de Cálculo I, pois ela só é oferecida no 2º período letivo para um dos cursos; nem as notas do Enem, porque a UFAM adota o SiSU

(Sistema de Seleção Unificada) para 50% das vagas da graduação presencial. A outra metade é preenchida por processo seletivo próprio (Processo Seletivo Contínuo - PSC), aplicado de forma seriada e contínua nos três anos do ensino médio. Como os professores responsáveis por IP e AL1 não foram os mesmos de 2012 a 2017 e, em alguns anos, os ingressantes foram divididos em turmas com professores distintos, esse fator também não foi considerado neste estudo.

### **3.2.2 Classificação do desempenho no 1º período com base no questionário CSED**

Após verificados quais atributos da Etapa 1 são melhores estimadores para classificar a situação final do aluno no curso (formado ou desistente), utilizou-se um algoritmo de árvore de regressão para verificar quais indicadores coletados pelo questionário CSED estão mais associados à evasão.

Árvores de regressão são utilizadas para particionar uma variável de resposta contínua em subconjuntos, com base na sua relação com uma ou mais variáveis explanatórias, que podem ser categóricas ou contínuas [Sutton 2005]. Métodos clássicos como a regressão linear tentam ajustar um modelo definido pelo analista para todo o espaço amostral. Já as árvores de regressão tentam particionar o espaço amostral em subgrupos, ajustando diferentes modelos para cada um deles [Venables e Ripley 2003]. Além disso, árvores de regressão não se baseiam em pressupostos de normalidade a respeito dos dados analisados. Por isso, são especialmente úteis quando há um grande número de variáveis envolvidas, pois podem ser usadas para identificar as variáveis e interações mais importantes [Sutton 2005].

## **4. Resultados e Análise**

Nesta seção, são apresentados os resultados obtidos nas duas etapas do estudo, bem como as análises correspondentes.

### **4.1. Etapa 1: Classificação da situação final do aluno com base no histórico**

Do ponto de vista da coordenação de um curso de Graduação, o sucesso acadêmico corresponde à situação final do estudante no curso: formado ou desistente. Outros indicadores podem ser utilizados, mas o cerne da discussão sobre sucesso acadêmico se resumirá, invariavelmente, à situação que o estudante atingiu no final do seu percurso acadêmico: formou ou desistiu?

Portanto, chegou-se a um problema de classificação binária, cuja classe alvo é “desistente”, pois os alunos que forem classificados nela deverão receber um tratamento específico por parte da coordenação de curso.

#### **4.1.1. Caracterização da amostra**

Nesta etapa, foi analisado o histórico de 463 estudantes dos cursos de Ciência da Computação, Engenharia de Computação e Engenharia de Software, com a situação final de “formado” ou “desistente”, que ingressaram na UFAM entre 2012 e 2017. Destes, 88 estudantes (19,0%) eram mulheres e 375 (81,0%) eram homens, como mostra o gráfico da Figura 1. É interessante notar que, embora as mulheres sejam minoria entre os estudantes de computação, elas apresentam uma maior taxa de formação (proporção entre as barras de “Formados” e “Desistentes”).

Com respeito às políticas afirmativas, 348 (75,2%) estudantes ingressaram nas vagas de ampla concorrência, 41 (8,8%) nas vagas dependentes de renda, 38 (8,2%) nas vagas independentes de renda, e 36 (7,8%) em outras vagas.

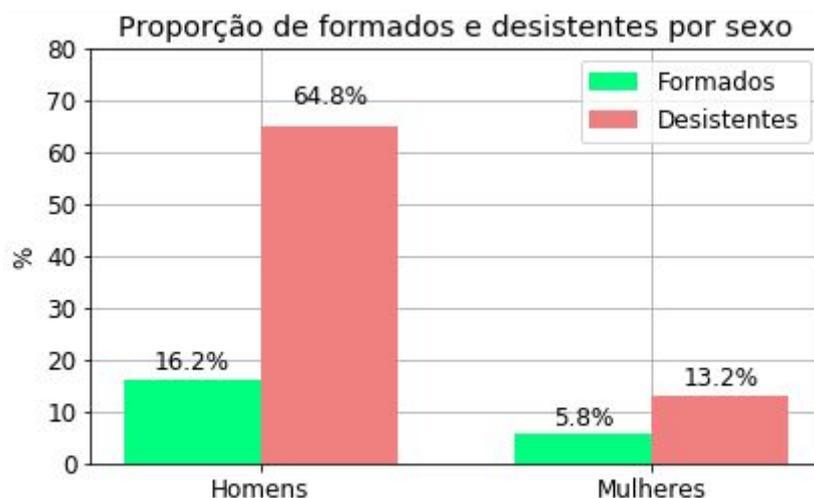


Figura 1. Distribuição por sexo dos alunos com situação final “formado” ou “desistente”, ingressantes em cursos de Computação de 2012 a 2017

#### 4.1.2. Seleção de indicadores de 1º período

Para cada registro de matrícula que atendesse às três condições descritas na Seção 3.2.1, foram coletados os seguintes indicadores: Sexo; Processo Seletivo; Cota; Coeficiente de rendimento no 1º período letivo (CR-1P); Taxa de aprovação em disciplinas do 1º período (TX-AP-1P); Nota em Álgebra Linear I (NotaAL1); Nota em Introdução à Programação (NotaIP).

Nesta etapa do estudo, três técnicas foram utilizadas para classificar os dados do histórico escolar: regressão logística, floresta aleatória (*random forest*) e *support vector machine* (SVM). Para validar os modelos, ou seja, para particionar os dados em conjunto de treinamento e conjunto de validação, utilizou-se o método *leave-one-out* [Witten *et al.* 2016]. O conjunto de atributos que resultou em uma melhor matriz de confusão foi CR-1P, TX-AP-1P e cota. A Tabela 1 apresenta as matrizes de confusão resultantes de cada técnica.

Verificou-se que o desempenho dos três classificadores foi muito semelhante. Outro resultado notável foi que todos os classificadores apresentaram maior erro de previsão para a categoria “desistente” do que para a de “formado” (26,5% contra cerca de 5%). Isso corrobora a observação de que alguns alunos, embora comecem o curso com baixo desempenho no 1º período, acabam acertando o prumo de sua carreira acadêmica nos demais períodos e logram em colar grau. Trata-se de uma informação baseada em evidências empíricas que as coordenações de curso podem usar para encorajar os estudantes com problemas de adaptação nos primeiros períodos dos cursos de graduação em computação e que desejam retomar seriamente os estudos.

**Tabela 1. Matrizes de confusão dos três classificadores utilizados para categorizar a situação final do aluno (formado e desistente)**

		Previsão						$\Sigma$
		Regressão Logística		Random Forest		SVM		
		Desistente	Formado	Desistente	Formado	Desistente	Formado	
Real	Desistente	94,5%	5,5%	93,6%	6,4%	91,7%	8,3%	<b>361</b>
	Formado	25,5%	72,5%	26,5%	73,5%	23,5%	76,5%	<b>102</b>
$\Sigma$		<b>369</b>	<b>94</b>	<b>365</b>	<b>98</b>	<b>355</b>	<b>108</b>	<b>463</b>

#### 4.2. Etapa 2: Associação dos indicadores do questionário CSED com o desempenho no 1º período

Na primeira etapa do estudo, verificou-se que os atributos passíveis de serem determinados no 1º período letivo e que melhor predizem a situação final do estudante são o coeficiente de rendimento no 1º período letivo (CR-1P) e a taxa de aprovação em disciplinas do 1º período (TX-AP-1P). Com base nisso, o objetivo da segunda etapa do estudo foi determinar que indicadores passíveis de serem identificados nas duas primeiras semanas de aula estão associados a esses dois atributos.

Para atingir esse objetivo, a partir de 2018, o questionário CSED mencionado na Seção 3.1 foi aplicado aos alunos ingressantes de três cursos de graduação presencial na área de Computação oferecidos pela UFAM no campus Manaus.

##### 4.2.1. Caracterização da amostra

Nesta segunda etapa do estudo, foram analisadas as respostas ao questionário CSED de 127 estudantes, entre 152 ingressantes nos cursos de Ciência da Computação, Engenharia de Computação e Engenharia de Software no ano de 2018. Entre os respondentes, 23 (18,1%) eram mulheres e 104 (81,9%) eram homens, distribuição similar à dos ingressantes entre 2012 e 2017 (etapa 1).

Com respeito às políticas afirmativas, 65 (51,2%) estudantes ingressaram nas vagas de ampla concorrência, 32 (25,2%) nas vagas dependentes de renda, e 30 (23,6%) nas vagas independentes de renda. Nenhum estudante ingressou em outro tipo de vaga no ano de 2018 nos três cursos analisados.

##### 4.2.2. Associação de indicadores socioeconômicos e demográficos com desempenho no 1º período letivo

Dos 38 itens do questionário CSED, 20 puderam ser mapeados em variáveis de resposta única, de escala categórica, aqui chamados de “indicadores”. Dado o pequeno número de amostras (127 respondentes em 2018), somente alguns indicadores deveriam ser selecionados como entrada para o modelo de árvore de regressão. Como todos os indicadores são de natureza categórica, não é possível estabelecer relação de correlação estatística deles com as variáveis categóricas. Por isso, adotamos o conceito de medida de associação estatística [Dancey e Reidy 2017], determinada pelo teste de t-student,

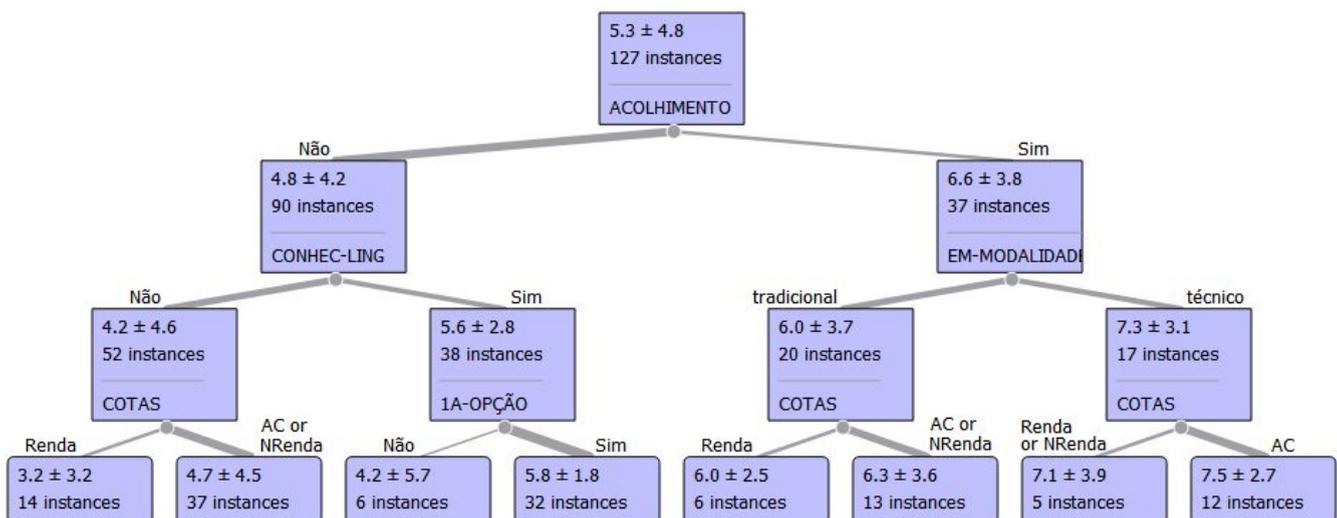
quando o atributo apresentava apenas duas categorias, ou pelo ANOVA, quando haviam três ou mais categorias no atributo.

**Tabela 2. Indicadores socioeconômicos do estudante e as respectivas significâncias de associação com variáveis de desempenho acadêmico no 1º período letivo: coeficiente de rendimento (CR-1P) e taxa de aprovação (TX-AP-1P)**

<b>Indicador</b>	<b>Categorias</b>	<b>valor p (CR-1P)</b>	<b>valor p (TX-AP-1P)</b>
<b>Curso</b>	<b>{CC, EC, ES}</b>	<b>0,005</b>	0,177
Sexo	{Feminino, Masculino}	0,188	0,330
Processo seletivo	{PSC, SiSU}	0,539	0,200
<b>Cota</b>	<b>{AC, Renda, IndepRenda}</b>	<b>0,027</b>	<b>0,009</b>
<b>Curso foi 1ª opção?</b>	<b>{Sim, Não}</b>	<b>0,022</b>	<b>0,006</b>
<b>Participação do acolhimento</b>	<b>{Sim, Não}</b>	<b>0,000</b>	<b>0,000</b>
<b>Graduação Prévia</b>	<b>{Sim, Não}</b>	0,555	<b>0,014</b>
<b>Estabelecimento EM</b>	<b>{Público, Privado}</b>	0,073	<b>0,035</b>
<b>Modalidade EM</b>	<b>{Tradicional, Técnico}</b>	<b>0,000</b>	<b>0,000</b>
<b>Conhecimento prévio de programação</b>	<b>{Sim, Não}</b>	<b>0,000</b>	<b>0,012</b>
Tipo de deslocamento	{Ônibus, Carona, Veículo próprio}	0,365	0,134
Tempo de deslocamento	{<1h, 1h-2h, >2h}	0,987	0,666
Estado Civil	{Solteiro, Casado}	0,443	0,191
<b>Filhos</b>	<b>{Sim, Não}</b>	0,171	<b>0,039</b>
Trabalho	{Não trabalha, Estágio, Tempo integral, Tempo parcial, Informal}	0,697	0,187
Computador em casa	{Sim, Não}	0,283	0,463
Uso exclusivo do computador	{Exclusivo, Compartilhado}	0,134	0,178
Acesso a internet	{Sim, Não}	0,721	0,759
Moradia	{Própria, Alugada}	0,240	0,488
Renda	{<1,5SM; 1,5SM a 3SM; 3SM a 4,5SM; 4,5SM a 6SM; >6SM}	0,588	0,377

Como critério de seleção, adotamos a significância da associação de cada indicador com o coeficiente de rendimento no 1º período letivo (CR-1P) e com a taxa de aprovação em disciplinas do 1º período (TX-AP-1P). A Tabela 2 apresenta o valor  $p$  das associações encontradas e destaca os indicadores em que a correlação é significativa, ou seja, apresentam valor  $p < 0,05$ .

Com base na Tabela 2, foram selecionados cinco indicadores como entrada para a árvore de regressão: cota, curso como 1ª opção, participação nas atividades de acolhimento, modalidade de Ensino Médio, e conhecimento prévio de programação. A variável curso, que teve uma associação significativa com o CR-1P, mas não apresentou associação significativa com a TX-AP-1P, não foi selecionada para o modelo de árvore de regressão, pois foi entendida como uma meta-informação sobre o estudante.



**Figura 2. Árvore de regressão para o coeficiente de rendimento de 1º período (CR-1P)**

A partir da árvore de regressão exibida na Figura 2, percebe-se que o indicador “participação do acolhimento” tem o maior impacto sobre o coeficiente de rendimento de 1º período (CR-1P). Entre os 37 participantes, o CR médio foi maior do que entre os 90 não participantes.

A acolhida é uma atividade de adesão voluntária promovida pelo centro acadêmico de Ciência da Computação, duas semanas antes do início oficial das aulas na UFAM. Os alunos veteranos se revezam ministrando aulas de introdução à programação e de revisão de matemática do Ensino Médio, contando com apoio institucional para reserva de salas e laboratórios. Nesse período, o calouro conhece os setores da universidade, as rotas dos ônibus, locais para comer, entre outros detalhes, sem a pressão das aulas da graduação. Como o ingresso na UFAM se desenvolve em diversas convocações de matrícula, que se estendem para além do início das aulas, os participantes da acolhida costumam ser os calouros que se matricularam nas duas primeiras chamadas.

Algumas hipóteses para explicar a relação entre a atividade de acolhida e o CR no 1º período letivo são as seguintes:

1. A adaptação paulatina com o ambiente universitário, promovida pelo acolhimento, oferece aos calouros condições para um melhor desempenho ao longo do 1º período letivo.
2. Os alunos que participam da acolhida são aqueles que obtiveram melhor desempenho nos processos seletivos adotados pela UFAM. Logo, o CR seria impactado não pela atividade de acolhida em si, mas sim pelo conhecimento prévio trazido pelos calouros, expresso em termos de rendimento nos processos seletivos e, conseqüentemente, pela ordem de preenchimento das vagas.

Infelizmente, os dados colhidos neste estudo não são suficientes para confirmar ou refutar as duas hipóteses acima, o que fica para um estudo futuro.

No segundo nível da árvore de regressão, estão os indicadores de modalidade do Ensino Médio (tradicional ou técnico) e o conhecimento prévio de linguagem de programação. Nota-se que a experiência prévia em programação, seja por curiosidade ou pelo contato com o ensino técnico formal, tem impacto sobre o CR do 1º período.

Em seguida, nota-se que os calouros ingressantes nas vagas de cota dependente de renda (renda familiar bruta per capita igual ou inferior a 1,5 salário mínimo que tenha cursado integralmente o ensino médio em escolas públicas) são os que obtiveram os menores coeficientes de rendimento no 1º período de 2018. Algumas explicações possíveis para esse fato são: a oportunidade de entrada não é acompanhada de políticas para estimular a permanência desses estudantes, e a vulnerabilidade socioeconômica faz com que tenham que dividir o tempo de estudo com outras tarefas pessoais ligadas à própria subsistência e da família.

Por fim, a escolha do curso como primeira opção no processo seletivo também impacta, ainda que em menor grau, o valor do CR, confirmando a observação do senso-comum.

## **6. Conclusão e Trabalhos Futuros**

A alta evasão de estudantes em cursos de graduação presencial na área de Computação é um problema grave, pois indica falhas no processo de formação, que se traduzem em custos financeiros e pessoais a todos os envolvidos. Suas causas são complexas, pois não é possível isolar todos os fatores envolvidos, que vão desde a seleção dos candidatos, até as metodologias de ensino e aprendizagem e de gestão dos cursos.

Neste trabalho, verificou-se que o desempenho acadêmico no 1º período letivo, medido em termos de coeficiente de rendimento e de taxa de aprovação em disciplinas, é um bom estimador da situação final do estudante: formado ou desistente. Como esses atributos somente são conhecidos após o final do 1º período letivo, partiu-se para a análise de indicadores passíveis de serem levantados durante as duas primeiras semanas de aula.

Verificou-se que a participação de atividade de acolhimento ao calouro foi o principal indicador a impactar o coeficiente de rendimento no 1º período letivo, seguido

da experiência prévia do estudante com linguagens de programação, seja de modo informal, ou por meio do Ensino Médio na modalidade de ensino técnico.

Como trabalho futuro, é necessário separar e entender que fatores intrínsecos à atividade de acolhimento estão de fato impactando o coeficiente de rendimento no 1º período letivo. Também encontra-se em andamento a análise das respostas dos ingressantes 2019 ao questionário CSED, o qual deseja-se comparar com o rendimento acadêmico no 1º período de 2019, a fim de confirmar ou não as tendências observadas em 2018. Por fim, estamos aplicando outro questionário junto aos estudantes que ingressaram antes de 2018, para levantar quais fatores que mais impactam seu desempenho acadêmico, bem como em que intensidade eles percebem esse impacto.

## Referências

- Azevedo, L. A., Santos, Y. S. (2015) Mineração de dados aplicada ao estudo da evasão e desempenho dos alunos do bacharelado em Ciência da Computação. Trabalho de Conclusão de Curso. Universidade de Brasília.
- Dancey, C. P. e Reidy, J. (2017). *Statistics without maths for psychology*, 7th. Pearson Education.
- Günther, H. (2003). Como elaborar um questionário. Série: Planejamento de pesquisa nas ciências sociais, (01).
- Hoed, R. M. (2016). Análise da evasão em cursos superiores: o caso da evasão em cursos superiores da área de Computação. Dissertação de mestrado. Brasília, DF: Universidade de Brasília.
- Manhães, L. M. B., Cruz, S. M. S., Costa, R. J. M., Zavaleta, J., e Zimbrão, G. (2011). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 1, No. 1).
- Matos, R. S., Souza, I. M., Silva, F. M., Beck, G. G. L. (2019). Gestão do Planejamento e Acompanhamento de Atividades Docentes de uma Universidade Federal. *Revista GUAL*. Florianópolis, v. 12, n. 1, pp.118-137.
- Morães, M. J. F., & Pombeiro, O. J. (2016). Evasão nos cursos de graduação em Computação de Curitiba. *Anais do EVINCI-UniBrasil*, 1(4), 2088-2103.
- Rigo, S. J., Cazella, S. C., & Cambruzzi, W. (2012). Minerando Dados Educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. In *Anais do Workshop de Desafios da Computação Aplicada à Educação* (pp. 168-177).
- Silva, F. L. (2001) Reflexões sobre o conceito e a função da universidade pública. *Estudos Avançados*, vol.15, n.42, pp.295-304.
- Sutton, C. D. (2005) Classification and regression trees, bagging, and boosting. In Rao, C. R. (Org.). *Handbook of Statistics*, vol. 24. Amsterdã: Elsevier, p. 303-329.
- Witten, I. H., Frank, E., Hall, M. A., Pal, C. J. (2011). *Data Mining: Practical machine learning tools and techniques*, 4th ed. Morgan Kaufmann.