

# Analizando a influência de atributos demográficos no desempenho de estudantes em uma 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}@icomp.ufam.edu.br

**Resumo.** *O baixo desempenho de alunos em disciplinas introdutórias de programação é cenário comum em cursos de graduação de ciências exatas e de engenharia. Fatores externos ao ambiente acadêmico são tão decisivos em relação aos resultados quanto os aspectos inerentes ao processo de aprendizagem. Com base na aplicação de técnicas de mineração de dados educacionais sobre os dados de uma instituição de ensino superior brasileira, observou-se que o desempenho em uma disciplina de introdução à programação para non-majors está correlacionado a características socioeconômicas e educacionais prévias ou a um subconjunto criado a partir da combinação destes atributos. Por sua vez, identificar antecipadamente estudantes em risco de reprovação pode ajudar a reduzir a evasão em disciplinas de programação.*

**Abstract.** *The low student performance in CS0 courses is a common scenario in undergraduate programs in science and engineering. External factors to the academic environment are as decisive in relation of these students results as the inherent aspects of the learning process. Based on the application of educational data mining techniques on the data of a Brazilian higher education institution, the student performance in a CS0 course for non-majors was observed to be correlated to previous socioeconomic and educational characteristics or to a subset created from the combination of these attributes. In its turn, the early identification of students at risk of failure can help to mitigate dropout in CS0 courses.*

## 1. Introdução

O desempenho insuficiente de alunos de graduação em disciplinas de introdução à programação, conhecidas na literatura como *Computer Science 0* (CS0), tem motivado a realização de diversas pesquisas no campo da mineração de dados educacionais a fim de se identificar as causas desse fenômeno [Martins, Lopes e Raabe 2012].

Bosse e Gerosa (2015) apontam, por exemplo, com base em levantamentos realizados na Universidade de São Paulo, que o índice de reprovação e trancamento em CS0 é por volta de 29%. Além disso, outra parcela considerável de alunos finaliza tais disciplinas com obtenção de conhecimento de maneira inconsistente, sendo poucos aqueles que obtêm desempenho de excelência.

Uma abordagem individualizada por parte dos docentes, dedicando atenção e acompanhamento específico a cada aluno visando mitigar as dificuldades de aprendizado, torna-se inviável pelo fato dessas turmas de CS0 serem habitualmente numerosas [Khosravi e Cooper 2017].

Antecipar as probabilidades de desempenho do estudante pode levar à condução de ações pedagógicas mais eficientes em sala de aula e à aplicação de melhoria dos planos e métodos de ensino vigentes. Além disso, intervenções institucionais mais efetivas no combate à evasão e reprovação em disciplinas de programação [Ahadi *et al.* 2015] poderão ser postas em prática com o auxílio do resultado desta pesquisa.

Este estudo tem como objetivo construir uma caracterização dos dados socioeconômicos coletados de alunos de cursos de graduação da área de ciências exatas e de engenharia em disciplinas introdutórias de programação. Para isso, foram coletados dados de três períodos letivos de 14 cursos dessas áreas na Universidade Federal do Amazonas (UFAM). Pretende-se, ainda, disponibilizar uma visão numérica inicial sobre os dados coletados e possibilitar a tomada de decisão quanto aos próximos passos para a construção de um modelo de predição de desempenho estudantil em tais disciplinas.

O presente artigo está estruturado em torno das seguintes questões de pesquisa:

1. **(Q1)** Quais dos atributos coletados a partir de informações sociodemográficas e do cadastro acadêmico dos alunos têm maior correlação com a média final em CS0?
2. **(Q2)** Existe correlação entre camadas distintas de dados (sociais, econômicos, familiares, conceitos prévios, etc.) de atributos de estudantes em relação ao desempenho acadêmico em CS0?

O artigo está organizado da seguinte forma: na Seção 2, estão descritos os trabalhos relacionados a esta pesquisa e que serviram como referencial a mesma; a Seção 3 descreve a metodologia de avaliação aplicada; na Seção 4, estão descritos os dados e o contexto educacional da pesquisa e os resultados discutidos na Seção 5; uma análise sobre tais resultados é apresentada na Seção 6 e, por fim, a Seção 7 apresenta as conclusões e trabalhos futuros.

## **2. Trabalhos relacionados**

Possíveis vinculações de atributos demográficos, sociais e comportamentais ao baixo desempenho de alunos em disciplinas introdutórias de programação têm objeto de pesquisas científicas desde a década de 1950 [Leinonen *et al.* 2016].

Tais trabalhos vêm sendo elaborados objetivando prever o desempenho de estudantes em disciplinas de programação [Pascoal, Brito e Rêgo 2015], [Byrne 2001] ou, ainda, a necessidade de suporte pedagógico a estes, a fim de que sejam evitadas reprovações e evasões [Martins, Lopes e Raabe 2012]. Em tais pesquisas foram apontados fatores determinantes de sucesso ou insucesso no ensino de programação.

Alguns trabalhos afirmam, por exemplo, que possuir conhecimento prévio em programação tem um efeito positivo na obtenção de sucesso em disciplinas introdutórias de programação [Byrne e Lyons 2001], [Hagan e Markham 2000].

Técnicas de agrupamento têm sido usadas para, nos casos de salas de aula numerosas, categorizar os programadores iniciantes usando análise automatizada de registros para revelar padrões na programação dos estudantes [Khosravi e Cooper 2017].

Ademais, em outros trabalhos, métodos de seleção de atributos já foram aplicados na previsão da insucesso de graduandos em disciplinas, inclusive, com evasão do curso [Júnior; Kaestner e Noronha 2016]. Nesse caso, os experimentos foram realizados com informações extraídas do sistema acadêmico da instituição de ensino superior, sendo selecionados atributos tais como sexo, estado civil e nota da redação de alunos dos cursos presenciais de graduação com oferta semestral, ingressantes a partir do exame nacional do ensino médio (ENEM).

Outros trabalhos baseados em *learning analytics* esforçaram-se por revelar relações e padrões relevantes entre as notas dos alunos e suas interações com o ambiente de aprendizagem [Ahadi *et al.* 2016], inclusive a linguagem de programação escolhida com foco na previsão do desempenho acadêmico [Frag *et al.* 2013].

Esses dados vêm sendo utilizados tanto na própria área de *learning analytics* quanto na mineração de dados educacionais (*educational data mining*), a fim de que os tais informações sejam empregadas na predição do desempenho de discentes, usando técnicas de análises estatísticas, mineração de dados e aprendizagem de máquina, de modo que não se avalie tão somente o produto enviado pelo aluno, mas as circunstâncias por trás deste também [Denny *et al.* 2010].

Ahadi *et al.* (2015) utilizaram dados demográficos como atributos em algoritmo de aprendizagem de máquina com o intuito de alavancar a acurácia dos modelos preditivos. Nesse experimento, atributos como gênero, desempenho acadêmico, experiência anterior em programação e até comportamento em laboratório foram utilizados na tentativa de otimizar a identificação antecipada da performance de discentes em CS0. Contudo, a maioria dos resultados encontrados não trouxeram correlação estatísticas significantes.

Um outro modelo para identificação precoce de alunos propensos ao insucesso nas disciplinas de programação em cursos de Computação foi construído a partir da utilização de informações de desempenho dos alunos nas provas do vestibular e as suas notas nas disciplinas de programação cursadas na faculdade [Pascoal *et al.* 2015].

Em todos os casos, o objetivo principal do estudo foi investigar a viabilidade da obtenção destas informações em etapas iniciais de realização do curso [Hagan e Markham 2000], [Frag, Sanwar e Debzani 2013], para apoiar a tomada de ações antecipadas por parte dos docentes e, até mesmo, da instituição [Bosse e Gerosa 2015], restando, contudo, a necessidade de conclusões mais acuradas a respeito da seleção dos atributos mais relevantes no decurso de processos de predição de resultados, conforme pretendido através das questões de pesquisa levantadas neste trabalho.

### **3. Metodologia aplicada**

A pesquisa em questão tem por objetivo identificar as eventuais correlações entre o desempenho acadêmico de estudantes das disciplinas CS0 e os dados demográficos coletados destes alunos a partir da extração de dados de atributos de estudantes dos cursos de graduação da área de ciências exatas e de engenharia, visando subsidiar a adoção de estratégias institucionais preventivas de combate à reprovação e evasão de discentes.

Com base em técnicas de mineração de dados educacionais, foram cruzadas informações proveniente de duas bases de dados educacionais: histórico de atividades de um sistema juiz online e registros de discentes no sistema de controle acadêmico. Em

seguida, foram definidos os atributos de um aluno a serem 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.

Com a utilização de dados oriundos do questionário demográfico existente no sistema Juiz Online utilizado na pesquisa, a priori, foram selecionados nove atributos encontrados na base de dados da aplicação citada, conforme Tabela 1.

**Tabela 1: Tabela geral de atributos selecionados [Fonte própria]**

Item	Derivado	Atributo	Valores
A01	Não	Sexo	Masculino ou feminino
A02	Não	Estado_civil	Solteiro, casado e viúvo
A03	Não	Experiencia_programacao	Experiência prévia em alguma linguagem de programação (sim ou não)
A04	Sim	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
A05	Não	Filhos	Possuir ou não filhos
A06	Não	Origem_ensino_medio	Público convencional – privado convencional – médio técnico
A07	Não	Origem_vaga	Ampla concorrência - Cota Independente de renda - Cota com renda baixa - Outros
A08	Não	Internet_em_casa	Possuir ou não internet em casa
A09	Não	PC_exclusivo_em_casa	Possuir ou não PC de uso exclusivo em casa

#### 4. Descrição dos dados e do contexto educacional

Neste estudo, foram coletados dados anonimizados de estudantes de 14 cursos de graduação presencial nas áreas de engenharia e ciências exatas em que a disciplina de Introdução à Programação (CS0) era um componente curricular obrigatório. Os dados são provenientes de turmas realizadas em três semestres letivos consecutivos: o 1º e o 2º de 2017 e o 1º de 2018, totalizando 1.040 registros de matrícula, entre os quais 848 correspondiam a alunos únicos.

Do total de alunos coletados, 547 eram de alunos do sexo masculino (64,5%) e 301 eram de alunos do sexo feminino (35,5%). A Tabela 2 apresenta a idade média dos participantes, distribuída por sexo.

**Tabela 2: Média e desvio padrão da idade dos alunos estudados, por sexo [Fonte: Sistema de Informações para o Ensino/UFAM]**

Sexo	Idade	
	Média	Desvio Padrão
Feminino	20,5	2,9
Masculino	21,9	5,2
<b>Total</b>	<b>21,4</b>	<b>4,6</b>

Todos os docentes alocados a essas turmas seguiram uma mesma metodologia híbrida de ensino e aprendizagem, que mescla o ensino presencial com atividades virtuais baseadas em um juiz online. Segundo Wasik *et al.* (2018), juízes online são plataformas web cujo propósito é compilar, executar e avaliar programas submetidos por usuários. Por meio deste, foram realizadas avaliações formativas (prática de exercícios de programação, com consulta à bibliografia, professores, tutores e colegas) e somativas (avaliação prática sem consulta) [Carvalho, Oliveira e Gadelha 2016].

A Tabela 3 apresenta uma síntese sobre as informações inerentes aos dados coletados durante a pesquisa, delimitando, dentre outras questões, a fonte das informações coletadas, o escopo e período avaliados.

**Tabela 3: Tabela síntese de informações sobre a pesquisa [Fonte própria]**

Item	Características
População-alvo	Alunos de graduação das áreas de engenharia e ciências exatas
Limitação de escopo	Disciplinas introdutórias de programação (CS0)
Fonte de dados	Tabelas do juiz online e do sistema de controle acadêmico
Período avaliado dos dados	janeiro/2017 a julho/2018
Quantidade de períodos letivos	03 (três)
Quantidade total de registros de matrícula	1.040
Cursos de graduação	14 cursos de ciências exatas e engenharia (non-majors)
Atributos selecionados somados aos novos criados (derivados)	09 atributos
Variável dependente	Situação final na disciplina (aprovado ou reprovado)

## 5. Análise e discussão sobre os resultados

Na fase exploratória do processo de mineração de dados educacionais com testes estatísticos de hipóteses foram aplicadas técnicas de análise de dados que objetivaram identificar se os dados obtidos obedeciam a uma distribuição normal.

Por sua vez dizer que uma amostra segue uma distribuição normal é verificar se a maioria dos valores dessa amostra está próxima ao valor médio encontrado nela própria.

Para realização da análise normalidade da amostra foi aplicado o Teste de Shapiro-Wilk utilizando-se o software R. A partir dos dados coletados, verificou-se que todas as amostras seguem uma distribuição normal.

Os nove atributos coletados (Tabela 1) foram comparados ao atributo de saída correspondente ao Resultado de cada aluno em CS0: aprovado ou reprovado.

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

### 5.1 Atributos coletados com maior correlação com a média final em CS0

Através da aplicação do teste exato de Fisher e de medidas de associação baseadas no coeficiente de contingência proposto por Pearson (1904), verificou-se, conforme dados dispostos na Tabela 4, com dados ordenados decrescentemente por grau de intensidade de associação.

**Tabela 4: Resultados dos testes estatísticos de intensidade de associação entre atributos sociodemográficos e o resultado em CS0 [Fonte própria]**

Item	Atributo	Valor p	intensidade de associação
A06	origem_ensino_medio	0,353	0,19
A04	faixa_etaria	0,217	0,09
<b>A05</b>	<b>filhos</b>	<b>0,009</b>	<b>0,08</b>
<b>A01</b>	<b>sexo</b>	<b>0,022</b>	<b>0,07</b>
A02	estado_civil	0,487	0,06
A08	internet_em_casa	0,179	0,06
A09	pc_exclusivo_em_casa	0,129	0,05
A07	origem_vaga	0,689	0,03
A03	experiencia_programacao	0,355	0,02

Obtivemos valores entre 0 e 1 para a medida de associação entre os atributos analisados e o resultado de alunos nas disciplinas CS0. Frisa-se que é atingido o seu valor mínimo se, e somente se  $Q^2_{obs} = 0$ , ou seja, no caso de completa independência. Por sua vez, é atingido o seu valor máximo, e somente se  $Q^2_{obs} = 1$ , ou seja, no caso de completa dependência.

Os atributos observados, em sua totalidade obtiveram grau de intensidade de associação consideravelmente baixos, indicando baixo acoplamento aos desempenhos dos alunos. Em suma, os resultados apresentados na Tabela 4 fortalecem a tese da necessidade de combinação de atributos para derivação de terceiros na busca de novos atributos com maior intensidade de associação ao desempenho de alunos em CS0, conforme descrito na seção 5.2.

## **5.2 Correlação entre atributos de estudantes e o desempenho acadêmico em CS0**

Antecipa-se que a média geral de aprovação encontrada na amostra foi de 39%, ficando, consequentemente, a média de reprovação em torno de 61% dos discentes inscritos nos dois anos averiguados: 2017 e 2018.

Estudos anteriores muitas vezes apontaram o sexo dos estudantes como um dos fatores que podem explicar a aptidão de programação. Uma das razões pode ser que o campo da computação é às vezes visto como sendo dominado por homens, e, portanto, apresenta uma cultura orientada para o masculino.

Entretanto, os resultados não mostraram nenhuma correlação significativa vinculada ao sexo do estudante e o resultado em CS0.

Nos dados analisados das turmas de 2018, 35% dos alunos matriculados em CS0 eram do sexo feminino e 65% eram do sexo masculino. Dentre os homens, 66% não possuíam experiência prévia em programação e eram originários do ensino médio público convencional. Esse grupo obteve um índice de aprovação de 43%.

Analisando exclusivamente o quantitativo de reprovados, os índices para os demais atributos já discutidos são bastante similares. Chama atenção, contudo, a predominância de alunos solteiros (97%), sem filhos (96%), com acesso a computador pessoal (85%) e internet em casa (70%) dentre aqueles sem sucesso na disciplina.

Do total de mulheres, 37% obtiveram êxito nas disciplinas e foram aprovadas. Ao se combinar esse atributo com a origem do ensino médio técnico, obtivemos uma melhora relevante nos índices encontrados. Passamos para um quantitativo de 70% dessas mulheres com aprovação em CS0, quase o dobro do inicial.

O referido grupo apresenta, ainda, um índice elevado de 61% de estudantes com experiência em programação, enquanto que o índice anterior que levava em consideração somente o sexo feminino foi de apenas 22%, praticamente um terço índice resultado da combinação dos atributos.

Percebe-se, desta forma, que a combinação dos atributos origem\_ensino\_medio (técnico) com o sexo feminino e a experiencia\_programacao eleva consideravelmente as chances de aprovação em disciplinas introdutórias de programação. Já para os alunos do sexo masculino, a origem do ensino médio proveniente do tipo técnico não foi tão impactante quanto para o sexo feminino, mais ainda assim correspondeu a uma melhora de 19% no indicador de aprovação, passando de 43% para 49% a incidência de sucesso.

Nos dados das turmas avaliadas de 2018, dentre o total de alunos investigados que possuíam filhos (6%) e eram casados (5%), foi observado um índice de 76% de reprovação nas disciplinas introdutórias de programação. Por outro lado, a ausência de filhos e cônjuge para estudantes do sexo feminino não interferiu significativamente os índices de aprovação, elevando os percentuais de sucesso em apenas 1%.

Na totalidade dos casos acima, foi observado que, além de ter acesso a computador em casa, em 88% das situações identificou-se a incidência de acesso à internet concomitantemente. Este fato demonstra que esses dois últimos atributos citados não influenciaram, neste caso, no desempenho em CS0 e, portanto não correspondem a classificadores eficientes para predição de resultados nessas disciplinas.

Contudo, pôde ser observado nesse mesmo grupo de reprovados que nenhum dos alunos possuía origem do ensino médio técnico. Do total de alunos reprovados neste conjunto, 94% advinha de escolas públicas convencionais e 6% de escolas particulares.

Os piores resultados encontrados foram para os alunos de origem do ensino médio público convencional, com apenas 36% de aprovação. Já para os alunos originados do ensino médio particular convencional, o índice de aprovação foi de 48%.

Em se tratando exclusivamente de alunos de origem do ensino médio técnico (13% do total investigado para 2018), 55% destes obtiveram aprovação na disciplina de CS0, índice consideravelmente superior à média geral que foi de 39% para o ano de 2018.

Os resultados obtidos identificaram, ainda, que 67% dos integrantes deste grupo possuíam algum tipo de experiência anterior em linguagem de programação. Quanto maior a idade, maiores foram os índices registrados de insucesso, ainda que a quantidade de alunos diminua com o incremento da faixa etária.

Além disso, a análise sobre os dados coletados demonstrou que os atributos estado civil, filhos, experiência profissional anterior e em programação estão vinculados de maneira diretamente proporcional à faixa etária. Quanto maior a idade, maior também são os índices positivos dos referidos atributos, sem, contudo, modificar positivamente os percentuais de aprovação nas disciplinas de CS0.

A exceção encontrada foi a combinação com a origem do ensino médio, caso no qual a situação de alunos oriundos de escolas técnicas gerou um incremento de 40% no percentual de sucesso na disciplina, passando de 40% para 56% de aprovação.

A condição de existência ou não de computador pessoal em casa não incorreu em influência significativa aos resultados de sucesso e insucesso nas disciplinas de programação, ainda que tenha havido uma diferença de 16% entre a condição de possuir e não possuir um PC em casa.

A origem da vaga de admissão dos alunos, sendo esta cotista (por renda ou independentemente desta) ou ampla concorrência, não gerou influência nos resultados de aprovação ou reprovação nas disciplinas de introdução à programação. Os valores encontrados para este atributo foram extremamente próximos, tanto para os índices de aprovações quanto os de fracasso nas disciplinas.

## **6. Conclusões e trabalhos futuros**

O desempenho acadêmico de estudantes em CS0, como em qualquer outra disciplina, pode estar eventualmente influenciado, compreensivelmente, a fatores extracurriculares.

Os resultados obtidos nesta pesquisa demonstraram que, na maioria dos casos, características ligadas à dimensão educacional, tais como origem do ensino médio e experiência com linguagens de programação influenciam os resultados obtidos em CS0 de maneira muito mais efetiva que qualquer outro atributo ligado à dimensão social do estudante como sexo, estado civil ou existência de filhos.

Desta forma, o atributo `origem_ensino_medio` é passível de consideração destacada no processo de predição de resultados de aprovação ou reprovação de discentes em CS0. Em suma, a pesquisa demonstrou que a totalidade de alunos com origem do ensino médio técnico obteve aprovação em CS0, mesmo tendo sido observado uma incidência de 73% destes alunos já possuírem filhos dentre os alunos já casados, por exemplo.

O atributo `experiencia_em_programacao` também foi identificado como uma potente característica a ser levada em consideração, isoladamente ou em conjunto.

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 preambular de substituição de uma análise, não esporadicamente tangencial, encontrada em boa parte das ações direcionadas à melhoria dos resultados acadêmicos dos alunos de disciplinas introdutórias de programação de computadores dos cursos de graduação das áreas de engenharia e ciências exatas.



Por esses motivos, sustenta-se a intenção de aprofundamento desta pesquisa e desenvolvimento de um modelo preditivo de desempenho de estudantes de CS0, com base em atributos socioeconômicos, a fim de que, entre outras questões, possam ser abordadas políticas institucionais de apoio a estudantes com risco de abandono da disciplina ou evasão escolar.

Como trabalhos futuros, pode-se coletar e analisar um número maior de atributos socioeconômicos, bem como selecionar aqueles mais relevantes para a construção de um modelo probabilístico de predição de desempenho acadêmico em CS0.

## Referências

- Ahadi, A. e Lister, R. (2015) “Exploring Machine Learning Methods to Automatically Identify Students in Need of Assistance.” In ACM Technical Symposium on Computing Science Education, SIGCSE '16. Omaha, Nebraska, USA: ACM.
- Ahadi, A., Lister, R. e Vihavainen, A. (2016) “On the Number of Attempts Students Made on Some Online Programming Exercises During Semester and Their Subsequent Performance on Final Exam Questions.” Pp. 218–23 in *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education, ITiCSE '16*. New York, NY, USA: ACM.
- Bosse, Y. e Gerosa, M. A. (2015) “Reprovações e Trancamentos nas Disciplinas de Introdução à Programação da Universidade de São Paulo: Um Estudo Preliminar”. In XXIII Workshop sobre Educação em Computação (WEI). Recife.
- Byrne, P. e Lyons, G. (2001). “The effect of student attributes on success in programming”. In ACM SIGCSE Bulletin, volume 33, pages 49–52. ACM.
- Carvalho, L., Oliveira, D. e Gadelha, B. (2016) “Juiz online como ferramenta de apoio a uma metodologia de ensino híbrido em programação.” In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação (SBIE)).
- Hagan, D. e Markham, S. (2000) “Does it help to have some programming experience before beginning a computing degree program?” ACM SIGCSE Bulletin, 32(3):25–28.
- Farag, W., Sanwar, A. e Debzani, D. (2013) “Does Language Choice Influence the Effectiveness of Online Introductory Programming Courses?” Pp. 165–70 in Proceedings of the 14th Annual ACM SIGITE Conference on Information Technology Education, SIGITE '13. New York, NY, USA: ACM.
- Júnior, J., Noronha, R. e Kaestner, C. (2017) “Método de Seleção de Atributos Aplicados na Previsão da Evasão de Cursos de Graduação”. Revista de Informática Aplicada, Volume 13, Número 2.
- Khosravi, H. e Cooper, K. (2017) “Using Learning Analytics to Investigate Patterns of Performance and Engagement in Large Classes” . In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education. New York, NY, USA: ACM.
- Leinonen, J., Longi, K., Klami, A. e Vihavainen, A. (2016) “Automatic Inference of Programming Performance and Experience from Typing Patterns.” Pp. 132–37 in

Proceedings of the 47th ACM Technical Symposium on Computing Science Education, SIGCSE '16. New York, NY, USA: ACM.

Martins, L., Lopes, D. e Raabe A. (2012) “Um Assistente de Predição de Evasão Aplicado a Uma Disciplina Introdutória do Curso de Ciência da Computação.” Anais do Simpósio Brasileiro de Informática na Educação 23(1):26–30.

Pascoal, T., Brito, D. e Rêgo, T. (2015) “Uma abordagem para a previsão de Aplicado a uma Disciplina Introdutória do Curso de Ciência da Computação.” Anais do Simpósio Brasileiro de Informática na Educação 23(1):26–30.

Wasik, S., Antczak, M., Laskowski, A. e Sternal, T. (2018) “A Survey on Online Judge Systems and Their Applications”. ACM Computing Surveys, 51(1), 3.