

# Um modelo bayesiano que auxilia na identificação de alunos com dificuldades na aprendizagem de programação de computadores

Willys S. A. Campos<sup>1</sup>, Rodrigo Quites Reis<sup>2</sup>, Otávio Noura Teixeira<sup>3</sup>

<sup>1</sup> Faculdade Estácio do Pará – Belém, PA – Brasil

<sup>2</sup> Laboratório de Engenharia de Software  
Universidade Federal do Pará – Belém, PA – Brasil

<sup>3</sup> Laboratório de Computação Natural – Campus Universitário de Tucuruí  
Universidade Federal do Pará – Tucuruí, PA – Brasil

<sup>1</sup> willyscampos@gmail.com, <sup>2</sup> quites@ufpa.br, <sup>3</sup> onoura@gmail.com

**Abstract.** *The learning of computer programming subjects has always brought challenges to any class of computer students. Generally, teachers who work in these subjects have signs about which students will become good programmers, but it is difficult to detect which ones need help in the learning process. This article proposes the use of a Bayesian model that helps in the identification, for teachers, of students with difficulties in the computer programming subjects. The results indicate potentiality in the contribution of student learning.*

**Resumo.** *O aprendizado das disciplinas de programação de computadores sempre trouxe desafios para qualquer turma de alunos de Computação. Geralmente, os professores que atuam nestas disciplinas têm sinais acerca de quais alunos se tornarão bons programadores, porém é difícil detectar quais necessitam de ajuda no processo de aprendizagem. Este artigo, propõe o uso de um modelo bayesiano que auxilia na identificação, para os professores, de alunos com dificuldades nas disciplinas de programação de computadores. Os resultados indicam potencialidade na contribuição da aprendizagem do aluno.*

## 1. Introdução

A Sociedade Brasileira de Computação (1999) reputa que, para uma carreira na área de Computação, as disciplinas de programação computacional são essenciais. Mas, muitos alunos apresentam dificuldades nesta aprendizagem. Entre os problemas relatados, destacam-se: a baixa compreensão das noções básicas de programação e o pensamento algorítmico segundo Yang et al. (2014). Nesse cenário, um modelo bayesiano, operacionalizado através de aprendizagem supervisionada, foi desenvolvido.

### 1.1 Motivação

A motivação para este trabalho é fortemente baseada na 8ª edição do Mapa do Ensino Superior - SEMESP (2018) que ressalta uma preocupação com os altos índices de evasão escolar vistas no primeiro ano de faculdade e nas habilidades dos profissionais do futuro que deverão ter conhecimento acerca de programação de computadores.

## 1.2 Objetivo

Considerando o contexto do presente trabalho, busca-se validar a hipótese de pesquisa “desenvolver um modelo bayesiano que auxilia na identificação de alunos com dificuldades na aprendizagem de programação de computadores”, usando uma abordagem de aprendizagem bayesiana.

## 1.3 Metodologia

O método utilizado no desenvolvimento deste trabalho foi o indutivo, partindo de práticas operacionais encadeadas da seguinte forma: criação de um problema, indicação de uma hipótese, coleta de dados e análise da resposta Lakatos e Marconi (2003). Assim, o trabalho seguiu com as etapas: uma pesquisa informal da literatura existente sobre *machine learning* e modelos bayesianos empregados no domínio de ambientes educacionais, reuniões com especialistas (Tabela 1); coleta de informações acadêmicas, com base no ano de 2018, com uma amostra de oitenta e sete alunos pertencentes ao curso de Sistemas de Informação. Ocorrendo, logo em seguida, o treinamento supervisionado do modelo, para posterior comparação das avaliações realizadas pelo modelo e pelos especialistas. O experimento em questão foi realizado em uma Faculdade em Belém do Pará.

**Tabela 1. Participantes especialistas da construção do modelo computacional**

Especialidade	Titulação	Participação	
		Avaliação dos Alunos	Construção do Modelo
Prof. Esp. 01	Msc. em Computação Aplicada	X	X
Prof. Esp. 02	Msc. em Computação Aplicada	X	X
Prof. Esp. 03	Msc. em Computação Aplicada	X	
Prof. Esp. 04	Esp. em Engenharia de Software	X	
Prof. Esp. 05	Esp. em Desenvolvimento de Sistemas WEB	X	X
Esp. Pedagógico 01	Coord. do Núcleo de Apoio Pedagógico		X

## 2. Fundamentação Teórica

Segundo Rich (1988), a Inteligência Artificial pode ser percebida como estudo de como fazer que os computadores realizem tarefas que atualmente os seres humanos são melhores. Para Russell e Norvig (2013) a análise preditiva das redes bayesianas baseia-se em conjuntos rotulados e prévios de informação histórica.

## 3. Trabalhos Correlatos

Todos os trabalhos correlatos identificados possuem aderência ao trabalho apresentado, devido a sempre usarem modelos de aprendizagem de máquina com abordagem de treinamento supervisionado e implementados em redes bayesianas. Todos, sem exceção, fazem uso de Inteligência Artificial (IA) para obtenção de respostas preditivas.

Seffrin et al. (2016) também usaram o modelo bayesiano para sugerir quanto um aluno possui de informação sobre o conhecimento algébrico. O principal diferencial desse artigo é a capacidade de avaliar, através de inferências, o conhecimento procedural e

conceitual, mostrando similaridades significativas entre a inferência da rede e os percentuais de acerto por parte dos estudantes. Esse trabalho, mostrou que existe a possibilidade de mapear outros tipos de conhecimento com o propósito de se chegar a uma conclusão sobre o grau de conhecimento do aluno, resultando em uma adequação aos mais variados tipos de alunos, analogamente ao artigo aqui apresentado.

O artigo de Longhi et al. (2010) usa uma máquina de inferência na teoria bayesiana para identificar os estados de ânimo de aluno. Com respostas de animado, desanimado, satisfeito ou insatisfeito, o objetivo é capacitar ambientes virtuais de aprendizagem para identificar a motivação do aluno baseado no seu estado de ânimo. A inferência proposta considera os traços de personalidade, os fatores motivacionais obtidos através dos padrões de comportamento e a subjetividade afetiva identificada em textos disponibilizados na plataforma de comunicação dos ambientes. No trabalho apresentado, mostrou-se eficaz a descrição do problema através de uma avaliação qualitativa e quantitativa com as inferências dos alunos para adaptação de ambientes dessa natureza.

O artigo de Porfírio et al (2018) usa uma rede bayesiana dinâmica para inferir conhecimento durante o processo de aprendizagem de programação de computadores. Onde, usa-se modelos de aprendizes, baseado nas habilidades dos alunos com o objetivo de facilitar a visualização, tanto dos conceitos já adquiridos e quais lacunas deverão ser preenchidas. O uso da ferramenta se mostrou benéfico no acompanhamento simultâneo de muitos alunos por parte do professor.

#### 4. Proposta de Modelo de Aprendizagem

Segundo Taylor e Parsons (2011), em trabalho prático, de como melhorar o envolvimento do aluno no processo de aprendizagem foi descoberta a existência de fatores e condições que podem contribuir na distinção dos alunos que não estão aprendendo e dos que estão aprendendo. Com isso, algumas variáveis podem ser destacadas para análise do aluno, incluindo: condições socioeconômicas, desempenho escolar, nível de conectividade com a internet e até grau de esforço do aluno. Baseado nessas premissas e no levantamento junto aos especialistas, uma proposição de modelo bayesiano foi construída (Figura 2).

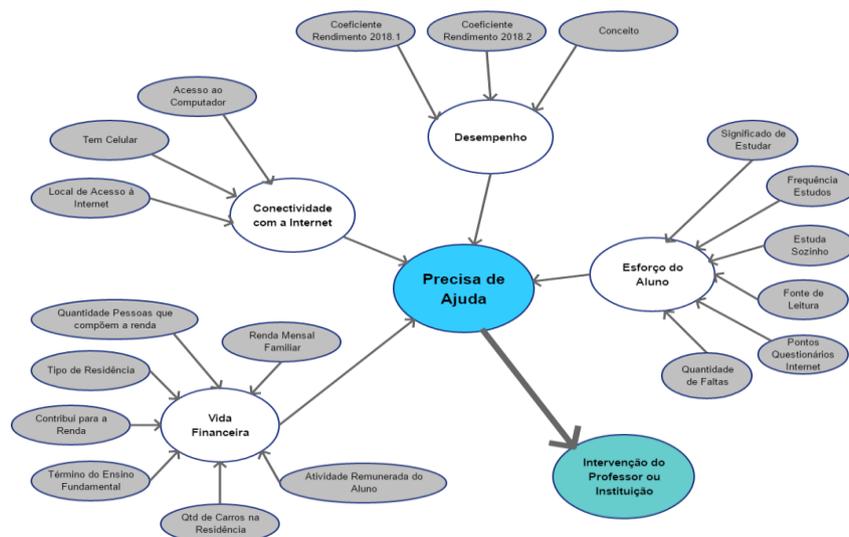


Figura 2. Modelo completo da rede bayesiana

No grafo acíclico direcionado (GAD) do modelo proposto (Figura 2), percebe-se graficamente a influência de todas as variáveis na obtenção da resposta esperada, que será dicotômica, SIM ou NÃO na intervenção ao aluno.

## 5. Implementação do Modelo

A linguagem Python, utilizada para implementação do trabalho, é uma linguagem de propósito geral que possui diversas aplicações no campo da Inteligência Artificial (IA).

No processamento das informações dos alunos, cerca de 90% destes dados foram utilizados no modelo, como base histórica de informação. Os passos para treinamento do modelo apresentado são: leitura da informação, pré-processamento dos atributos e treinamento do nó. A lista de todos os artefatos, utilizados no modelo, se encontra disponível em <<http://bit.ly/artefatos2019>>.

Na Tabela 2, encontra-se o comparativo entre os pareceres, no qual visualizam-se a avaliação de cada especialista (colunas identificadas como Aval 1 a Aval 5), o resultado obtido com o modelo proposto (coluna Modelo), e o conceito obtido pelo aluno ao final do período letivo de 2018 (Conceito final). Um dos avaliadores (Avaliador 4) recebeu apenas 8 fichas para emissão do seu parecer, devido a falha no processo humano de controle de entrega das fichas.

**Tabela 2. Comparativos de Respostas dos pareceres realizadas**

	<b>Aval 1</b>	<b>Aval 2</b>	<b>Aval 3</b>	<b>Aval 4</b>	<b>Aval 5</b>	<b>Modelo</b>	<b>Conceito Final</b>
<b>Aluno 1</b>	Não	Sim	Sim	-	Não	<b>Não</b>	BOM
<b>Aluno 2</b>	Não	Não	Não	Não	Não	<b>Não</b>	BOM
<b>Aluno 3</b>	Não	Não	Sim	Não	Não	<b>Não</b>	BOM
<b>Aluno 4</b>	Não	Não	Não	Sim	Não	<b>Não</b>	BOM
<b>Aluno 5</b>	Não	Sim	Não	Sim	Sim	<b>Sim</b>	REG
<b>Aluno 6</b>	Não	Não	Não	Sim	Não	<b>Não</b>	BOM
<b>Aluno 7</b>	Não	Sim	Sim	Sim	Não	<b>Sim</b>	REG
<b>Aluno 8</b>	Não	Não	Não	Não	Não	<b>Não</b>	BOM
<b>Aluno 9</b>	Sim	Sim	Sim	Sim	Não	<b>Sim</b>	REG

Percebe-se que em todos os resultados obtidos pelos avaliadores houve aderência do resultado obtido pelo modelo em relação a maioria das respostas dos especialistas. Na avaliação do Aluno 1, onde não houve posição definida pelos avaliadores, faltando o parecer do Avaliador 4, o modelo acabou indicando que o aluno não precisaria de apoio, o que foi confirmado pelo conceito final obtido pelo aluno.

## 6. Considerações Finais

Neste trabalho foi desenvolvido um modelo computacional bayesiano que usa o Teorema de Bayes como base na identificação de alunos com dificuldades em disciplinas que

envolvem programação de computadores. Os resultados, apresentam que o modelo de aprendizagem pode ser usado na abstração de uma determinada realidade e que pode ajudar positivamente o ambiente educacional no que tange ao caminho do aluno, inclusive influenciando semestres futuros e, conseqüentemente sua carreira profissional. A evolução do modelo está fortemente relacionada com vários aspectos identificados no trabalho. Foi percebido, na prática, que o incremento da base histórica tende a melhorar a acurácia, sendo esta, uma característica intrínseca nos modelos de aprendizagem supervisionada.

No estudo realizado com os especialistas (pedagogos e professores de programação) obteve-se como resultado o atual modelo construído neste trabalho. Mas, estas não são verdades absolutas e devem evoluir. Por exemplo, durante o desenvolvimento do trabalho, especificamente no treinamento de máquina inferido ao modelo, percebeu-se que todos os alunos de uma maneira ou de outra possuem acesso à Internet, deixando esse nó com baixa relevância devido a sua resposta de 100%. O modelo de aprendizagem de máquina baseia-se em reconhecimento de padrões. Logo, o estudo de novas disciplinas, inclusive de outras áreas do conhecimento, a identificação de aspectos regionais e especificidades da instituição podem ser levados em consideração para trabalhos futuros.

## Referências

- LAKATOS, E; MARCONI, A. **Metodologia científica**. São Paulo: Atlas, 2003.
- LONGHI, M; BEHAR, P; BERCHT, M. **Inferência dos estados de ânimo do aluno em um ambiente virtual de aprendizagem baseada em redes bayesianas**. Informática na educação: teoria & prática. 14. 10.22456/1982-1654.21935, 2011.
- PORFIRIO, A., PEREIRA, R., and MASCHIO, E. **Inferência de Conhecimento a Partir da Detecção Automática de Evidências no Domínio da Programação de Computadores**. Anais do XXIX Simpósio Brasileiro de Informática na Educação, 2018.
- RICH, E. **Inteligência Artificial**. São Paulo: McGraw-Hill, 1988.
- RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial**. Rio de Janeiro: Campus, 2013.
- SBC - Sociedade Brasileira de Computação. **Currículo de referência da SBC para cursos de graduação em computação e informática**. 1999.
- SEFFRIN, H; RUBI, G.; JAQUES, P. **Uma Rede Bayesiana aplicada à Modelagem do Conhecimento Algébrico do Aprendiz**. 10.5753/CBIE.SBIE.2013.597.
- SEMESP - Sindicato das Entidades Mantenedoras de Estabelecimentos de Ensino Superior no Estado de São Paulo. **Mapa do Ensino Superior**. São Paulo, 2018.
- TAYLOR, L.; PARSONS, J. **Improving Student Engagement**. *Current Issues in Education*, 14(1), 2011.
- YANG, T. C., YANG, S. J., & HWANG, G. J. **Development of an interactive test system for students' improving learning outcomes in a computer programming course**. In *Advanced Learning Technologies (ICALT)*, 2014 IEEE 14th International Conference on (pp. 637-639). IEEE. doi: 10.1109/ICALT.2014.186, 2014.