

Avaliação de Aprendizagem de *Machine Learning* na Educação Básica: Um Mapeamento da Literatura

Gustavo de Castro Salvador¹, Christiane Gresse von Wangenheim¹, Marcelo Fernando Rauber^{1,2}, Abisague Belem Garcia¹, Adriano F. Borgatto¹

¹Departamento de Informática e Estatística
Universidade Federal de Santa Catarina (UFSC) – Florianópolis – SC – Brasil

²Instituto Federal Catarinense (IFC) - *campus* Camboriú – SC – Brasil

gustavo.castro.salvador@grad.ufsc.br, c.wangenheim@ufsc.br,
marcelo.rauber@ifc.edu.br, abisague.garcia@posgrad.ufsc.br,
adriano.borgatto@ufsc.br

Abstract. *Machine learning (ML) is being increasingly integrated into our life. Thus, to demystify ML several instructional units are emerging teaching ML already in K-12. Yet, especially the assessment of the learning of ML concepts remains an open issue. Thus, in order to provide an overview on the current state of the art regarding the assessment of the learning of ML in K-12, we performed a systematic mapping. We identified 15 instructional units on ML, which also present the assessment of the students' learning, mostly in a simple manner by few test questions/quizzes with few proposing performance-based assessments. However, an evident lack of systematic definition and validation of these assessments indicates the need for further research.*

Resumo. *Machine Learning (ML) está cada vez mais integrada em nossa vida. Assim, para desmistificar ML, várias unidades instrucionais estão emergindo ensinando ML já na Educação Básica. No entanto, especialmente a avaliação da aprendizagem dos conceitos de ML continua sendo uma questão aberta. Assim, a fim de fornecer uma visão geral sobre o atual estado da arte em relação à avaliação da aprendizagem de ML na Educação Básica, realizamos um mapeamento sistemático. Identificamos 15 unidades instrucionais abordando ML, que também apresentam a avaliação da aprendizagem dos estudantes, principalmente de uma maneira simples como testes/questionários com poucos propondo uma avaliação baseada em desempenho. No entanto, há uma evidente falta de definição sistemática e validação dessas avaliações o que indica a necessidade de pesquisas futuras.*

1. Introdução

Machine Learning (ML) ou Aprendizado de Máquina tornou-se parte de nossa vida cotidiana, impactando profundamente nossa sociedade ao servir como motor de inovação para uma ampla gama de aplicativos, como sistemas de reconhecimento de voz, assistentes inteligentes, carros autônomos, etc. No entanto, a maioria das pessoas não entende a tecnologia por trás disso, o que pode tornar o ML um assunto misterioso ou até assustador, ofuscando seu potencial de impactar positivamente a sociedade [Evangelista *et al.*, 2018][Ho e Scadding, 2019].

Assim, para desmistificar o *Machine Learning* é importante introduzir conceitos e práticas básicas já na escola, permitindo que os estudantes se tornem não apenas consumidores, mas também criadores de soluções inteligentes [Touretzky *et al.*, 2019][Kandlhofer *et al.*, 2016]. A exposição a este conteúdo complexo pode melhorar as habilidades do dia a dia dos estudantes, equipando-os com conhecimento para lidar com as questões, sociais, econômicas e éticas criadas pelo uso de ML nas diversas esferas da sociedade [Kahn *et al.*, 2020]. Também pode encorajar estudantes a considerarem carreiras em Computação e fornecer uma preparação sólida para o ensino superior [CSTA, 2011]. O *Machine Learning*, como um subcampo da Computação, tem sido tradicionalmente ensinado apenas no ensino superior [Torrey, 2012][McGovern *et al.*, 2011]. E, embora hoje existam muitas iniciativas que têm como foco o ensino de programação e robótica, o ensino de Inteligência Artificial e ML para estudantes da Educação Básica está muito ausente [Hubwieser *et al.*, 2015].

No entanto, recentemente, algumas iniciativas e projetos que buscam ensinar ML na Educação Básica surgiram. Isto inclui diretrizes curriculares, p.ex., pela AI4K12 [Touretzky *et al.*, 2019] e unidades instrucionais, principalmente extracurriculares [Marques *et al.*, 2020]. Essas unidades instrucionais ensinam competências que vão de apresentações sobre o que é ML a técnicas específicas da área, com ênfase em redes neurais artificiais e impactos do ML até a aplicação e criação de modelos de *Deep Learning*. Poucas unidades instrucionais abordam o processo de ML de forma mais completa, com a maioria abordando somente uma parte, como preparação de dados, ou apresentação de alguns processos apenas de forma abstrata, por exemplo, o treinamento de redes neurais [Marques *et al.*, 2020]. Tipicamente são adotadas ferramentas visuais e/ou ambientes textuais como Jupyter Notebooks, às vezes de forma integrada com ambientes de programação baseado em blocos, como Scratch, SNAP! ou App Inventor [Gresse von Wangenheim *et al.*, 2020]. As unidades instrucionais variam também de unidades escolares e cursos no modo presencial até cursos e tutoriais *on-line*.

Como parte do processo de ensino, é importante avaliar a aprendizagem dos estudantes fornecendo *feedback* tanto para o aluno quanto para o professor [Hattie e Timperley, 2007]. No entanto, apesar de pesquisas existentes para avaliação do ensino de pensamento computacional e programação na Educação Básica [Tang *et al.*, 2019][Lye e Koh, 2014][Oliveira e Oliveira, 2014] observa-se a falta de propostas de avaliações da aprendizagem de ML do aluno [Marques *et al.*, 2020] como revisões sistemáticas, como p.ex. Camada e Durães (2020), Sanusi e Oyelere (2020), Zhou *et al.* (2020) não analisam de forma explícita esta questão. Portanto, o presente artigo visa levantar o estado da arte referente a avaliações da aprendizagem de ML de estudantes da Educação Básica por meio de um mapeamento sistemático da literatura.

2. Definição e execução do estudo de mapeamento sistemático

A fim de levantar o estado da arte é realizado um mapeamento sistemático da literatura seguindo o procedimento definido por Petersen *et al.* (2015).

2.1. Definição do Protocolo de Revisão

Pergunta de pesquisa. Quais modelos existem para avaliação de aprendizagem de *Machine Learning* do aluno na Educação Básica? Esta questão é refinada nas seguintes perguntas de análise:

AQ1. Quais unidades instrucionais voltadas ao ensino de ML na Educação Básica apresentam avaliações de aprendizagem?

AQ2. Quais são as características destas avaliações em termos de nível de aprendizagem, conteúdo e tipo?

AQ3. Que *feedback* instrucional é apresentado?

Critérios de seleção: Conforme o foco da pesquisa, são selecionados artigos relevantes seguindo os critérios de inclusão e exclusão apresentados no Quadro 1.

Quadro 1. Critérios de inclusão e exclusão utilizados nesta pesquisa.

Critérios de inclusão	Critérios de exclusão
<ul style="list-style-type: none">- Artigos escritos na língua inglesa.- Artigos que apresentem avaliações de aprendizagem de <i>Machine Learning</i> no contexto da Educação Básica.- Artigos publicados nos últimos dez anos de 2011 a 2021, considerando a tendência recente de ensino de conceitos de ML.	<ul style="list-style-type: none">- Artigos que enfocam o ensino de <i>ML</i> no Ensino Infantil e/ou Ensino Superior e/ou abordam Inteligência Artificial sem abordar conceitos de ML.- Artigos que somente apresentam uma avaliação do curso, mas não apresentam propostas para a avaliação da aprendizagem do aluno.- Publicações como blogs, vídeos ou outras ferramentas que não compõem uma unidade instrucional/curso para estudantes de Ensino Educação Básica.- Artigos que não apresentam informações substanciais, como artigos-resumo ou apenas de uma página.

Fontes. A busca é feita nos principais repositórios digitais no campo da Computação, incluindo Scopus, IEEE Xplore, ACM, SpringerLink, ScienceDirect, arXiv, SocArXiv, Google Scholar e MIT Media Lab. Além disso, foram realizadas buscas no Google Scholar e Google para complementar a pesquisa, minimizando o risco da omissão [Piasecki *et al.*, 2018]. Buscou-se também publicações no repositório do MIT Media Lab pela atuação nesta área.

String de busca. Com base na pergunta de pesquisa, foram realizadas pesquisas informais para calibrar a *string* de busca, identificando termos de pesquisa relevantes e sinônimos. Após obter poucos resultados relevantes com uma busca muito específica sobre avaliação em ML na Educação Básica, a estratégia de busca foi alterada visando em primeiro momento a busca por artigos que apresentassem unidades instrucionais voltadas ao ensino de ML, e em seguida considerados relevantes durante o processo de seleção somente artigos que incluíssem uma apresentação de avaliações. Para a busca de unidades instrucionais de ensino de ML foi definido a *string* seguindo Marques *et al.* (2020):

```
(teach* OR education OR course OR MOOC OR learn*) AND ("machine learning" OR "data science" OR "artificial intelligence" OR "deep learning") AND ("k-12" OR school* OR kids OR children OR teen*)
```

Observando também basicamente a inexistência de pesquisa voltada ao ensino de ML na Educação Básica em português, foi adotado somente um string de busca em inglês.

2.2. Execução da busca

A busca adaptada da *search string* a cada repositório foi realizada nos títulos e resumos limitada às áreas de Computação e Educação pelos autores em conjunto em março de 2021, conforme apresentado no Quadro 2.

Quadro 2. Número de artigos identificados por repositório e por fase de seleção

Fonte	No. de resultados da busca	No. de resultados analisados	No. de documentos potencialmente relevantes	No. de documentos relevantes
Scopus	3.760	500	37	0
IEEE Explore	1.113	500	11	1
ACM	4.151	500	25	4
SpringerLink	47.718	500	4	0
ScienceDirect	57	57	1	0
arXiv	140	140	4	3
SocArXiv	7.404	500	4	1
Google Scholar	17.000	500	11	3
Google	633.000	500	45	12
MIT Media Lab ¹	81	81	10	3
Total				15 (sem duplicatas)

A partir do resultado das buscas, foram selecionados artigos potencialmente relevantes de acordo com os critérios de inclusão e exclusão por meio de uma análise do título, resumo e palavra-chave de cada artigo. Foram analisados os primeiros 500 artigos mais relevantes encontrados em cada busca. Em seguida, foram analisados os artigos potencialmente relevantes pela sua leitura na íntegra. Muitos artigos se referem ao uso de técnicas de IA/ML na educação ao invés de ensinar ML. Estes foram excluídos, assim como artigos voltados a outros estágios educacionais e/ou artigos que apresentavam unidades instrucionais para ensinar ML, porém sem demonstrar algum tipo de avaliação da aprendizagem do aluno. Eliminando as duplicatas foram identificadas 15 unidades instrucionais relevantes como resultado final, conforme apresentado no Quadro 3. Como nem todas as unidades instrucionais disponibilizam explicitamente as informações desejadas, certas características foram inferidas pelos pesquisadores de acordo com a informação disponível nos próprios artigos e/ou cursos/plataformas educativas.

3. Resultados

3.1. Quais avaliações de aprendizagem de ML na Educação Básica existem?

Quadro 3. Artigos relevantes

ID	Nome	Descrição	Referência
[AI Family Challenge, 2019]	AI Family Challenge	Desafio para famílias aprenderem sobre IA e ML e resolver um problema em suas comunidades usando ferramentas de IA.	https://www.curiositymachine.org/about/
[Apps For Good, 2019]	Apps for Good: ML course	Curso que fornece uma visão geral de tópicos de ML e permite às equipes de estudantes projetar e construir um protótipo que resolva um problema real usando algoritmos de ML.	https://www.appsforgood.org/courses/machine-learning
[Brummelen <i>et al.</i> , 2020]	AI Literacy Workshop	Curso de IA a partir do qual os estudantes desenvolvem agentes de conversação usando uma interface no MIT App Inventor.	J. V. Brummelen <i>et al.</i> Teaching Tech to Talk: K-12 Conversational Artificial Intelligence Literacy Curriculum and Development Tools. <i>arXiv</i> , 2020.
[Code.org, 2019]	AI for Oceans	Curso <i>on-line</i> de introdução a ML envolvendo classificação de peixes do oceano como parte	https://curriculum.code.org/hoc/plugged/9/

¹<https://appinventor.mit.edu/explore/research>

		dos cursos da code.org.	
[CSER, 2020]	Teaching Artificial Intelligence in the Secondary Classroom	Curso gratuito de IA voltado para instrutores de IA usando ML em reconhecimento de imagens e visão computacional.	https://csermoocs.appspot.com/ai_secondary/
[Curiosity Machine, 2019]	Curiosity Machine - build a neural network	Desafio para construir um protótipo de uma rede neural artificial desconectada para classificação de diferentes objetos.	https://www.curiositymachine.org/challenges/126/
[Elements of AI, 2019]	Elements of AI	Curso <i>on-line</i> voltado ao ensino de ML para aprender o que é Inteligência Artificial, o que pode (e não pode) ser feito com IA e como começar a criar métodos de IA, inclusive ML.	https://course.elementsofai.com/4
[Exploring Computer Science, 2020]	Alternate Curriculum Unit: AI	Curso que leva os estudantes a explorarem aplicações práticas no dia a dia da IA e entender o impacto da IA em suas vidas por meio de diversas atividades.	http://www.exploringcs.org/wp-content/uploads/2019/09/AI-Unit-9-16-19.pdf
[Gresse von Wangenheim et al., 2020]	<i>Machine Learning</i> para Todos!	Curso que introduz conceitos básicos de ML, levando o aluno a criar um modelo de reconhecimento de imagens para separação de lixo reciclável usando o Google Teachable Machine.	C. Gresse von Wangenheim, et al. Machine Learning for All – Introducing Machine Learning in K-12. <i>SocArXiv</i> , 2020.
[Lee et al., 2021]	Developing Middle School Students' AI Literacy	Oficina de verão para preparar estudantes do Ensino Médio a desenvolver conhecimentos e habilidades fundamentais sobre IA e tornarem cidadãos informados e consumidores críticos da tecnologia.	Lee, I. et al. Developing Middle School Students' AI Literacy. <i>ACM Special Interest Group on Computer Science Education</i> , 2021.
[MIT App Inventor, 2019]	Introduction to Machine Learning: Image Classification	Curso que ensina noções básicas de ML e a criação de aplicativos que implementam esses conceitos por meio de classificação de imagens pelos estudantes com <i>deployment</i> via <i>apps</i> com App Inventor.	https://appinventor.mit.edu/explore/resources/ai/image-classification-look-extension
[ReadyAI, 2019]	ReadyAI AI + Me	Curso <i>on-line</i> destinado a fornecer aos estudantes os conceitos básicos de IA e de ML.	https://edu.readyai.org/courses/aime/
[Rodríguez-García et al., 2021]	LearningML	Oficina que ensina <i>Machine Learning</i> no contexto de desenvolvimento de projetos Scratch fornecendo uma plataforma web para o ensino de fundamentos de ML, permitindo treinar e testar modelos <i>on-line</i> e fazer o <i>deployment</i> via Scratch.	J. D. Rodríguez García et al. Evaluation of an Online Intervention to Teach Artificial Intelligence With LearningML to 10-16-Year-Old Students. <i>Technical Symposium of the ACM Special Interest Group on Computer Science Education</i> , 2021.
[Sakulkeakulsuk et al., 2018]	Kids making AI	Desafio de IA baseado em agricultura que visa ensinar o processo de criação de modelos de ML na forma de um jogo.	B. Sakulkeakulsuk et al. Kids making AI: Integrating Machine Learning, Gamification, and Social Context in STEM Education. <i>Proc. of IEEE Int. Conf. on Teaching, Assessment, and Learning for Engineering</i> , Wollongong, Australia, 2018.
[Tang, 2019]	Personal Image Classifier	Curso que ensina noções básicas de ML e faz com que os estudantes criem seus próprios aplicativos que implementam esses conceitos por meio de classificação de imagens com <i>deployment</i> via <i>apps</i> com App Inventor.	D. Tang. Empowering Novices to Understand and Use Machine Learning With Personalized Image Classification Models, Intuitive Analysis Tools, and MIT App Inventor. Master thesis, MIT, USA, 2019.

Foram encontradas 15 unidades instrucionais relevantes (Quadro 4) que apresentam algum tipo de avaliação de aprendizagem de ML voltado a estudantes da Educação Básica. Observa-se que mesmo tendo uma quantidade considerável de cursos

de ML para o Educação Básica (Marques *et al.*, 2020)(Camada e Durães, 2020), poucos abordam de forma explícita a avaliação da aprendizagem do aluno. A maioria visa o ensino de ML especificamente ou como parte da Inteligência Artificial de forma geral ao nível da Educação Básica. Poucas são especificamente voltadas ao Ensino Fundamental ou Ensino Médio.

3.2. Quais são as características destas avaliações?

Quadro 4. Características das avaliações

ID	Método de avaliação	Níveis de aprendizagem do Bloom	Fases do "Use-Modify-Create"	Conceitos de ML avaliados
[AI Family Challenge, 2019]	Quiz com uma questão de múltipla escolha	Conhecimento, Compreensão, Aplicação	Use	Conceitos básicos de ML
[Apps For Good, 2019]	Questões de avaliação	Conhecimento, Compreensão, Aplicação	Use, Create	Conceitos básicos de ML, impacto de ML
[Brummelen <i>et al.</i> , 2020]	Questões de avaliação, avaliação baseado em desempenho de ideias	Conhecimento, Compreensão e Aplicação	Use, Modify, Create	Conceitos básicos de ML, impacto de ML
[Code.org, 2019]	Conclusão da tarefa	Conhecimento, Compreensão	----	Conceitos básicos de ML
[CSER, 2020]	Questões abertas/de redação	Conhecimento, Compreensão	Use	Conceitos básicos de ML
[Curiosity Machine, 2019]	Quiz com 3 questões de múltipla escolha	Conhecimento, Compreensão	----	Redes neurais
[Elements of AI, 2019]	Exercícios com 1-3 questões abertas/de redação ou de múltipla escolha	Conhecimento, Compreensão	----	Conceitos básicos de ML, impacto de ML, redes neurais
[Exploring Computer Science, 2020]	Rubricas com critérios com 8 - 14 para a avaliação das apresentações dos estudantes, questões de redação	Conhecimento, Compreensão	----	Conceitos básicos de ML, impacto de ML
[Gresse von Wangenheim <i>et al.</i> , 2020]	Quiz, questões verdadeiro/falso, arrastar e soltar, rubrica de 11 critérios para a avaliação dos artefatos do desenvolvimento de um modelo de ML	Conhecimento, Compreensão, Aplicação	Use	Conceitos básicos de ML, redes neurais, processo de ML, impacto de ML
[Lee <i>et al.</i> , 2021]	Quiz, questões verdadeiro/falso, questões abertas/de redação	Conhecimento, Compreensão	Use, Modify	Conceitos básicos de ML, redes neurais
[MIT App Inventor, 2019]	Teste com 3 questões de múltipla escolha	Conhecimento, Compreensão, Aplicação	Use	Redes neurais
[ReadyAI, 2019]	Quiz de pergunta única (múltipla escolha, arrastar e soltar, etc.)	Conhecimento, Compreensão	----	Conceitos básicos de ML, impacto de ML
[Rodríguez-García <i>et al.</i> , 2021]	Quiz com 14 questões, exercícios, questão aberta	Conhecimento, Compreensão, Aplicação	Use, Modify, Create	Conceitos básicos de ML, redes neurais
[Sakulkeakulsuk <i>et al.</i> , 2018]	Gamificação de pontos de acordo com a performance dos modelos de ML	Conhecimento, Compreensão	Use	Conceitos básicos de ML
[Tang, 2019]	Teste com 3 questões de múltipla escolha	Conhecimento, Compreensão, Aplicação	Use	Redes neurais

A maioria das avaliações são relativamente simples, referindo-se somente à aprendizagem nos níveis de conhecimento e compreensão, seguindo a Taxonomia de Bloom *et al.* (1965), com poucas que abordam também o nível de aplicação. Como o foco da maioria dos cursos no nível introdutório, a aprendizagem de conceitos básicos de ML é a mais avaliada, menos frequentemente avalia-se a aprendizagem de conceitos de redes neurais. Várias avaliações abordam questões éticas e o impacto de ML na sociedade.

Em alguns casos, as avaliações consistem somente em questões de resposta única no final das unidades instrucionais ou apenas monitoram a conclusão da tarefa [Code.org, 2019]. Algumas unidades avaliam as respostas dos estudantes a exercícios [Elements of AI, 2019]. Várias unidades instrucionais usam quizzes com diferentes tipos de questões (verdadeiro/falso, múltipla escolha, arrastar e soltar, etc.), incluindo por exemplo, [Gresse von Wangenheim *et al.*, 2020][ReadyAI, 2019][Rodríguez-García *et al.*, 2021], enquanto MIT App Inventor (2019) e Tang (2019) utilizam testes com 3 questões de múltipla escolha. Ao levar em consideração que a maioria das unidades instrucionais são usualmente oferecidas como atividades extracurriculares, estes tipos de avaliações simples são adequados para prevenir a desmotivação dos estudantes. Entretanto, a falta de avaliações mais rigorosas pode impedir um melhor suporte para a aprendizagem dos estudantes e o aprimoramento das unidades instrucionais.

Poucas unidades adotam uma avaliação baseada em desempenho definindo rubricas para a avaliação de apresentações dos projetos [Exploring Computer Science, 2020] ou rubricas para avaliar artefatos de modelos de ML criados pelos estudantes [Gresse von Wangenheim *et al.*, 2020]. Dado a natureza recente das unidades instrucionais de ML para a Educação Básica, a maioria foca na avaliação de resultados na fase de *Use* do ciclo "*Use-Modify-Create*". Há exceções, como as unidades propostas por Rodríguez-García *et al.* (2021) e Apps for Good (2019), levando os estudantes a criarem modelos personalizados de ML. Essa criação é feita adotando a estratégia de "ação computacional" [Tissenbaum *et al.*, 2019], que permite que os estudantes aprendam enquanto criam artefatos significativos que têm um impacto em suas vidas e comunidades que estejam inseridos.

3.3. Que *feedback* instrucional é apresentado?

Quadro 6. Tipos de *feedback* apresentados

ID	Tipo de <i>feedback</i>	Tipo de automação
[AI Family Challenge, 2019]	Correção de respostas	Correção automatizada de respostas
[Apps For Good, 2019]	Avaliação manual pelo instrutor	Não possui
[Brummelen <i>et al.</i> , 2020]	Avaliação manual pelo instrutor	Não possui
[Code.org, 2019]	Indicação de conclusão da tarefa sem análise de correção. <i>Feedback</i> para instrutores para monitoramento de uma turma	Grau de conclusão da tarefa
[CSER, 2020]	Comparação manual por parte do aluno de um modelo ideal fornecido pelo instrutor	Não possui
[Curiosity Machine, 2019]	Apresentação das respostas fornecidas	Não possui
[Elements of AI, 2019]	Correção de respostas de múltipla escolha, apresenta respostas de exemplo e revisão por pares para respostas de texto. Há certificado, e a quantidade de exercícios realizados é monitorada	Correção automatizada de respostas de múltipla escolha
[Exploring Computer Science, 2020]	Avaliação manual pelo instrutor indicando o total de pontos	Não possui
[Gresse von Wangenheim <i>et al.</i> , 2020]	Correção de respostas, acompanhamento de conclusão de tarefas	Correção automatizada de respostas dos quizzes, não possui automação para avaliação da rubrica
[Lee <i>et al.</i> , 2021]	Avaliação manual pelo instrutor	Não possui
[MIT App Inventor, 2019]	Avaliação manual pelo instrutor	Não possui
[ReadyAI, 2019]	Correção de respostas, acompanhamento de conclusão de tarefas	Correção automatizada de respostas

[Rodríguez-García <i>et al.</i> , 2021]	Autocorreção testando o modelo	Cálculo de precisão e desempenho do modelo criado
[Sakulkeakulsuk <i>et al.</i> , 2018]	Avaliação manual pelo instrutor	Não possui
[Tang, 2019]	Avaliação manual pelo instrutor	Não possui

O *feedback* instrucional, tipicamente, é limitado à indicação se as questões foram respondidas corretamente, sem maiores informações. Muitas avaliações não possuem algum tipo de automação para efetuar as avaliações, sendo executadas por instrutores das unidades instrucionais. Basicamente, somente quizzes são automatizados nas unidades *on-line*. Uma exceção é a avaliação apresentada por Rodríguez-García *et al.* (2021), que retorna a precisão e desempenho do modelo criado para a avaliação do aluno. A avaliação de Alternate Curriculum Unit: AI [Exploring Computer Science, 2020] utiliza uma escala de pontuação para apresentar o *feedback* ao estudante com base na soma ponderada dos itens da rubrica, sendo julgado por juízes que representam especialistas do domínio. Ao final, algumas unidades oferecem um certificado para completar o curso e para aumentar o engajamento [Code.org, 2019]. E Sakulkeakulsuk *et al.* (2018) adotam uma abordagem de gamificação para aumentar a motivação do aluno também no momento da avaliação.

4. Discussão

Observa-se como resultado desta revisão que atualmente a maioria das unidades instrucionais de ensino de *Machine Learning* para estudantes da Educação Básica não apresentam propostas para a avaliação de aprendizagem dos estudantes. Isto pode ser explicado pelo fato que o ensino de ML na Educação Básica ainda está emergente com um aumento de unidades instrucionais principalmente a partir dos últimos dois anos [Marques *et al.*, 2020]. Mesmo vários artigos apresentando avaliações da qualidade de curso por meio de pré/pós-testes, no quesito de avaliação de aprendizagem do aluno usam somente uma autoavaliação não medindo de forma mais confiável a aprendizagem.

O enfoque da maioria dos cursos no nível iniciante explica a ênfase em níveis de aprendizagem mais baixos segundo a Taxonomia de Bloom, com poucos que levam e avaliam a competência ao nível de aplicação de ML. Assim, a maioria das avaliações foca em conceitos básicos a serem avaliados por meio de *quizzes* e testes. Foram identificados poucos exemplos de rubricas de avaliação de desempenho que analisam artefatos criados pelo aluno ao aplicar conceitos de ML. Outro fator que impede a avaliação de um escopo maior é a típica curta duração destes cursos, muitas vezes de forma extracurricular. O tipo de *feedback* apresentado para os estudantes costuma ser simples, o qual indica se a tarefa foi concluída ou realizada corretamente. A maioria das avaliações propostas são realizadas manualmente pelos instrutores e/ou juízes. Identificou-se a automação de avaliações somente no caso de cursos *on-line* referente a *quizzes*.

Com base nestes resultados é evidente a necessidade de aprimoramento de modelos de avaliação do ensino de ML na Educação Básica, tanto em termos do seu desenvolvimento e validação sistemática quanto uma cobertura maior dos objetivos de aprendizagem e automação para preparar uma futura adoção mais ampla em escolas brasileiras.

Ameaças à validade. Mapeamentos sistemáticos podem sofrer do viés comum de que os resultados positivos têm maior probabilidade de serem publicados do que os negativos.

No entanto, foi considerado que os resultados, sejam positivos ou negativos, têm apenas uma pequena influência nesse mapeamento, uma vez que foi buscado caracterizar os modelos de avaliação de aprendizagem. A fim de mitigar o risco de omitir algum artigo relevante, a *string* de busca foi construída para ser o mais abrangente possível, considerando não apenas os principais conceitos, mas também sinônimos e a busca em diversas fontes. Ameaças relacionadas a seleção e extração de dados foram mitigadas por meio da definição detalhada dos critérios de inclusão e exclusão, além da revisão da seleção e extração em conjunto com todos os coautores até que um consenso fosse obtido.

5. Conclusão

O presente artigo expõe um mapeamento sistemático sobre abordagens existentes de avaliações de aprendizagem de ensino de *Machine Learning* no contexto da Educação Básica. Identificamos 15 unidades instrucionais em ML, que também apresentam a avaliação da aprendizagem dos estudantes, a maioria de uma maneira simples, com poucos testes/questionários e poucas propostas de avaliação baseada em desempenho. A maioria das avaliações tem enfoque em níveis introdutórios abrangendo somente conceitos básicos de ML e níveis mais baixos de aprendizagem da Taxonomia de Bloom. O *feedback* fornecido ao aluno também se restringe à indicação da resposta correta e conclusão da tarefa, sem indicação de informações mais construtivas que guiem o processo de aprendizagem do aluno. A maioria das avaliações também não oferece um suporte automatizado, limita-se basicamente a cursos *on-line* em relação a correção de respostas de *quizzes*. Estes resultados demonstram uma falta aparente de avaliações confiáveis e válidas de forma mais completa e abrangente, o que indica a necessidade de pesquisas nesta área para suportar a introdução do ensino de ML de forma mais ampla em escolas brasileiras levando em consideração a democratização desse conhecimento.

Agradecimentos

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil (CNPq).

Referências

- Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. Taxonomy of educational objectives. New York: David McKay Company, 1956.
- Camada, M. Y.; Durães, G. M. Ensino da Inteligência Artificial na Educação Básica: um novo horizonte para as pesquisas brasileiras. Anais do Simpósio Brasileiro de Informática na Educação, online, 2020.
- CSTA. CSTA K-12 Computer Science Standards. ACM, New York, NY, USA, 2011.
- Evangelista, I.; Blesio, G.; Benatti, E. Why Are We Not Teaching Machine Learning at High School? A Proposal. World Engineering Education Forum, Albuquerque, NM, USA.,2018.
- Galan, D.; Heradio, R.; Vargas, H.; Abad, I.; Cerrada, J. A. Automated Assessment of Computer Programming Practices: The 8-Years UNED Experience. IEEE Access, 7, 2019.
- Gresse von Wangenheim, C.; Hauck, J. C. R.; Pacheco, F. S.; Bertoni, M. F. Ferramentas Visuais para o Ensino de Machine Learning na Educação Básica. RENOTE, 18(2), 2020.

- Hattie, J; Timperley, H. The power of feedback. *Review of Educational Research*, 77(1), 2007.
- Ho, J. W. K.; Scadding, M. Classroom Activities for Teaching Artificial Intelligence to Primary School Students. *Proc. of the Int. Conference on Computational Thinking*, Hong Kong, China, 2019.
- Hubwieser, P. *et al.* A Global Snapshot of Computer Science Education in K-12 Schools. *Proc. of the ITiCSE on Working Group Reports*, Vilnius, Lithuania. 2015.
- Kahn, K. M.; Lu, Y., Zhang, J.; Winters, N.; Gao, M. Deep learning programming by all. *Proc. of the Conference on Constructionism*, Dublin, Ireland. 2020.
- Kandlhofer, M.; Steinbauer G.; Hirschmugl-Gaisch S.; Huber, P. Artificial Intelligence and Computer Science in Education. *Proc. of the Frontiers in Education Conference*, Erie, PA, USA, 2016.
- Lye, S. Y.; Koh, J. H. L. Review on teaching and learning of computational thinking through programming: What is next for K-12?. *Computers in Human Behavior*, 41, 2014.
- Marques, L. S.; Gresse von Wangenheim, C.; Hauck, J. C. R. Hauck. Ensino de Machine Learning na Educação Básica: um Mapeamento Sistemático do Estado da Arte. *Anais do Simpósio Brasileiro de Informática na Educação*, Natal, Brasil, 2020.
- McGovern, A.; Tidwell, Z.; Rushing, D. Teaching Introductory Artificial Intelligence through Java-Based Games. *Proc. of the 2nd Symposium on Educational Advances in Artificial Intelligence*, San Francisco, CA, USA, 2011.
- Oliveira, M.; Oliveira, E. Metodologia de Diagnóstico e Regulação de Componentes de Habilidades da Aprendizagem de Programação. *Anais do Workshop sobre Educação em Computação*, Brasília, Brasil, 2014.
- Petersen, K.; Feldt, R.; Mujtaba, S.; Mattsson, M. Systematic Mapping Studies in Software Engineering. *Proc. of the 12th Int. Conference on Evaluation and Assessment in Software Engineering*, Bari, Italy, 2008.
- Piasecki, J.; Waligora, M.; Dranseika, V. Google Search as an Additional Source in Systematic Reviews. *Science and Engineering Ethics* 24, 2018.
- Sanusi, I. T.; Oyelere, S. S. Pedagogies of Machine Learning in K-12 Context. *IEEE Frontiers in Education Conference* Uppsala, Sweden, 2020.
- Tang, D.; Utsumi, Y.; Lao, N. PIC: A Personal Image Classification Webtool for High School Students. *Proc. of the Int. Joint Conferences on Artificial Intelligence*, Macao, China, 2019.
- Tissenbaum, M.; Sheldon, J.; Abelson, H. From computational thinking to computational action. *Communications of the ACM*, 62(3), 2019.
- Torrey, L. Teaching Problem-Solving in Algorithms and AI. *Proc. of the 3rd Symposium on Educational Advances in Artificial Intelligence*, Toronto, Ontario, Canada, 2012.
- Touretzky, D. S. et al. Envisioning AI for K-12: What Should Every Child Know about AI? *Proc. of the 33rd Conf. on Artificial Intelligence*, Honolulu, HI, USA, 2019.
- Zhou, X.; Brummelen, J. V.; Lin, P. Designing AI Learning Experiences for K-12: Emerging Works, Future Opportunities and a Design Framework, *arXiv*, 2020.