



UNIVERSIDADE FEDERAL DE SANTA CATARINA

PROGRAMA DE PÓS-GRADUAÇÃO EM MÉTODOS E GESTÃO EM AVALIAÇÃO

Abisague Belem Garcia

**AVALIAÇÃO DE APRENDIZAGEM DE CONCEITOS DE *MACHINE LEARNING* NOS
ANOS FINAIS DO ENSINO FUNDAMENTAL E ENSINO MÉDIO**

Florianópolis

2022

Abisague Belem Garcia

**AVALIAÇÃO DE APRENDIZAGEM DE CONCEITOS DE *MACHINE LEARNING* NOS
ANOS FINAIS DO ENSINO FUNDAMENTAL E ENSINO MÉDIO**

Dissertação submetida ao Programa de Pós-Graduação em Métodos e Gestão em Avaliação da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Métodos e Gestão em Avaliação.

Orientador: Prof. Dr. Adriano Ferreti Borgatto

Coorientadora: Prof.^a Dr. rer. nat. Christiane Gresse von Wangenheim, PMP

Florianópolis

2022

Abisague Belem Garcia

**AVALIAÇÃO DE APRENDIZAGEM DE CONCEITOS DE *MACHINE LEARNING* NOS
ANOS FINAIS DO ENSINO FUNDAMENTAL E DO ENSINO MÉDIO**

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Dr. Adriano Ferreti Borgatto

Prof^a. Dra. Andreia Zanella

Prof. Dr. Jean Hauck

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Métodos e Gestão em Avaliação.

Coordenação do Programa de Pós-Graduação

Prof. Dr. Adriano Ferreti Borgatto

Orientador

Profa. Dra. rer. nat. Christiane Gresse von Wangenheim, PMP

Coorientadora

Florianópolis

2022

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter me concedido saúde no decorrer deste trabalho. Agradeço ao professor Adriano e à professora Christiane, pelas suas orientações, prestatividade, e apoio durante todo o período de desenvolvimento deste trabalho.

Agradeço aos professores Jean, Marcelo e Ramon pelas suas contribuições indispensáveis nas etapas deste trabalho.

Agradeço aos membros do GQS/UFSC (Grupo de Qualidade de Software) e da Iniciativa Computação na Escola que contribuíram efetivamente para a conclusão deste trabalho.

Agradeço aos professores que aceitaram participar como membros da banca, pelos seus comentários e sugestões enriquecedoras.

Agradeço ao meu esposo e filha, que sempre me incentivaram e me deram o suporte necessário para concluir este trabalho.

RESUMO

Tratando-se de computação, uma das principais áreas hoje é o *Machine Learning* (ML), ou aprendizagem de máquina, um campo de estudos em que algoritmos dão ao computador a habilidade de aprender com dados e se tornou parte da nossa vida cotidiana, impactando nossa sociedade. Para desmistificar o *Machine Learning* é importante introduzir conceitos e práticas já na Educação Básica, com o intuito de preparar os estudantes aos desafios do Século XXI. Nesse contexto, especialmente a avaliação da aprendizagem do ensino de ML é importante para acompanhar o progresso de aprendizagem do aluno, mas o ensino de competências de *Machine Learning* atualmente não é incluído na Educação Básica. A maioria das Unidades Instrucionais têm sido propostas como unidades extracurriculares, e não há modelos comumente adotados de como avaliar a aprendizagem de competências de *Machine Learning* na Educação Básica. Assim, o presente trabalho visa responder à pergunta de pesquisa: Como avaliar a aprendizagem de competências de *Machine Learning* dos alunos nos Anos Finais do Ensino Fundamental e do Ensino Médio? Desta forma o objetivo deste trabalho é desenvolver um modelo de avaliação da aprendizagem de competência de *Machine Learning* nos Anos Finais do Ensino Fundamental e do Ensino Médio. O desenvolvimento inclui tanto a revisão sistemática de literatura quanto a definição sistemática de um modelo de avaliação com base no desempenho do aluno voltadas ao ensino de *Machine Learning* aplicado à classificação de imagens usando o ambiente visual Google Teachable Machine. O modelo é instanciado por uma rubrica e avaliada estatisticamente em relação à sua confiabilidade e validade convergente, utilizando o Coeficiente Ômega e a Correlação Policórica. Com base nos dados de um total de 108 estudantes, coletados em 5 aplicações do ensino de ML, resultados da avaliação preliminar de confiabilidade e validade convergente indicam que a rubrica de avaliação do desempenho de aprendizagem de *Machine Learning* na Educação Básica está próxima de apresentar confiabilidade e validade convergente. Disponibilizando esse modelo de avaliação, espera-se facilitar a avaliação da aprendizagem de conceitos de *Machine Learning*, fornecendo suporte ao ensino de *Machine Learning* na Educação Básica.

Palavras-chave: Aprendizado de Máquina, Avaliação de Aprendizagem, Educação Básica, Revisão Sistemática, Coeficiente Ômega, Correlação Policórica

ABSTRACT

Abstract: When it comes to computing, one of the main areas today is Machine Learning (ML), a field of study in which algorithms give computers the ability to learn from data. ML has become part of our everyday life, impacting our society, and, thus in order to demystify Machine Learning it is important to introduce concepts and practices already in K-12, in order to prepare students for challenges of the 21st century. In this context, especially the assessment of the students' learning is important. But as the teaching of Machine Learning is currently not yet included in K-12 curricula, there are no commonly adopted models of how to assess the learning of Machine Learning in K-12. Therefore, the present work aims to answer the research question: How to assess the learning of the application of basic Machine Learning concepts by middle and high school students? The objective of this work is to develop a model for assessing the students' learning of Machine Learning at these educational stages. This includes a systematic literature review and the systematic definition of an assessment model with regard to student performance in the context of teaching image classification using Google Teachable Machine. The approach is instantiated by a rubric and statistically evaluated with respect to reliability and convergent validity, using the Omega Coefficient and the Polychoric Correlation. Based on data from a total of 108 students, collected in 5 applications in practice, results of the preliminary evaluation indicate that the rubric is close to presenting reliability and convergent validity. The result of this research, creating a performance-based assessment model, is expected to facilitate the assessment of the learning of Machine Learning concepts, in order to provide feedback and guide the students' learning process, thus, providing support for the teaching of Machine Learning in K-12.

Keywords: Machine Learning, Learning Assessment, Systematic Review, Omega Coefficient, Polychoric Correlation

LISTA DE FIGURAS

Figura 1 - Relação entre <i>Deep Learning</i> e a área de inteligência artificial.....	19
Figura 2 - Sistema de <i>Deep Learning</i> capaz de identificar o tipo de lixo a partir de uma imagem...20	
Figura 3 - Parâmetros de treinamento do Google Teachable Machine.....	22
Figura 4 - Taxa de Acurácia.....	23
Figura 5 - Matriz de confusão.....	23
Figura 6 - Precisão por época durante o treinamento.....	24
Figura 7 - Cinco grandes ideias em Inteligência Artificial.....	28
Figura 8 - Exemplos do material didático.....	36
Figura 9 - Visão geral das abordagens com avaliações e práticas de ML.....	46
Figura 10- Visão geral dos Níveis de aprendizagem.....	46
Figura 11 - Tipos de <i>feedbacks</i> apresentados nas Unidades Instrucionais.....	51
Figura 12 - Quantidade de <i>feedback</i> automatizado e manual.....	51
Figura 13 - Distribuição dos sistemas operacionais de celulares no Brasil em 2020.....	56
Figura 14 - Frequência de acesso à internet - Alunos do 9º Ano do Ensino Fundamental.....	57
Figura 15 - Dispositivos utilizados para acessar a internet.....	57
Figura 16 - Atividades com as TICs.....	58
Figura 17 - Alunos durante as aplicações do curso.....	64

LISTA DE QUADROS

Quadro 1 - Etapas da pesquisa.....	14
Quadro 2 - Como funciona o cérebro do computador - Rede Neural	18
Quadro 3 - Competências para avaliar as Tecnologias de IA.....	26
Quadro 4 - Rascunho do Conteúdo da grande ideia 3 - Progressão.....	29
Quadro 5 - Objetivos de aprendizagem do curso.....	32
Quadro 6 - Plano de ensino.....	34
Quadro 7 - Avaliação de aprendizagem.....	36
Quadro 8 - Possibilidades de atividades avaliativas de forma online.....	38
Quadro 9 - Palavras-chave.....	42
Quadro 10 - Strings de busca para cada fonte.....	42
Quadro 11 - Número de artigos identificados por repositório e por fase de seleção.....	43
Quadro 12 - Resumo das avaliações de conceitos e práticas de ML.....	44
Quadro 13 - Ano das publicações das Unidades Instrucionais.....	45
Quadro 14 - Exemplos dos tipos de avaliações.....	48
Quadro 15 - Exemplos de feedback automatizados e avaliação manuais.....	52
Quadro 16 - Rubrica de ML - Estágio de uso.....	60
Quadro 17 - Visão geral das inscrições do curso.....	64
Quadro 18 - Dados demográficos dos participantes.....	65
Quadro 19 - Visão geral dos dados coletados.....	65
Quadro 20 - Distribuição das respostas aos itens.....	66
Quadro 21 - Distribuição de respostas.....	67
Quadro 22 - Coeficiente Ômega - Análise com todos os itens.....	68
Quadro 23 - Coeficiente Ômega - Análise com os itens C1 a C5.....	68
Quadro 24 - Coeficiente Ômega - Análise com os itens C6 a C11.....	69
Quadro 25 - Matriz de correlação policórica dos itens.....	70
Quadro 26: Correlação dos pares de itens em cada dimensão.....	70

LISTA DE ABREVIATURAS E SIGLAS

AI4K12	<i>The Artificial Intelligence (AI) for K-12 initiative</i>
BNCC	Base Nacional Comum Curricular
CnE	Computação na Escola
CSTA	<i>Computer Science Teachers Association</i>
GQM	<i>Goal Question Metric</i>
IA	Inteligência Artificial
ML	<i>Machine Learning</i>
PPGMGA	Programa de Pós-Graduação em Métodos e Gestão em Avaliação
SI	Sala Informatizada
SBC	Sociedade Brasileira de Computação
TCLE	Termo de Consentimento Livre e Esclarecido
UFSC	Universidade Federal de Santa Catarina

SUMÁRIO

1 INTRODUÇÃO.....	11
1.1 CONTEXTUALIZAÇÃO.....	11
1.2 PROBLEMATIZAÇÃO	12
1.3 OBJETIVOS	14
1.4 METODOLOGIA DE PESQUISA.....	15
1.5 CONTRIBUIÇÕES.....	17
2 FUNDAMENTAÇÃO TEÓRICA	18
2.1 MACHINE LEARNING.....	18
2.2 ENSINO DE MACHINE LEARNING NA EDUCAÇÃO BÁSICA.....	26
2.3 CURSO DE MACHINE LEARNING PARA TODOS!.....	32
3 ESTADO DA ARTE	41
3.1 DEFINIÇÃO DO PROTOCOLO DE REVISÃO.....	41
3.2 EXECUÇÃO DA BUSCA.....	43
3.3 ANÁLISE DE DADOS	44
3.4 DISCUSSÃO	54
4 DESENVOLVIMENTO DO MODELO DE AVALIAÇÃO.....	57
4.2 DEFINIÇÃO DO PLANO DE AVALIAÇÃO	61
5 AVALIAÇÃO DO MODELO.....	63
5.1 DEFINIÇÃO DA AVALIAÇÃO.....	63
5.2 APLICAÇÃO E COLETA DE DADOS	64
5.3 PREPARAÇÃO DE DADOS	66
5.4 RESULTADOS DA ANÁLISE.....	68
5.4.1 Existe evidência de consistência interna da rubrica para a avaliação da aprendizagem de ML do estudante?.....	68
5.4.2 Existe evidência de validade convergente da rubrica?.....	70
5.4.3 Discussão	72
5.4.4 Ameaças à validade.....	73
6 CONCLUSÃO	74
REFERÊNCIAS BIBLIOGRÁFICAS.....	75
ANEXO - TERMOS E QUESTIONÁRIOS.....	80

1 INTRODUÇÃO

Este capítulo contextualiza a pesquisa, apresentando também o objetivo deste trabalho, o método de pesquisa utilizado para realizá-lo e as contribuições esperadas com a sua realização.

1.1 CONTEXTUALIZAÇÃO

Ensinar os fundamentos de Computação desde a Educação Básica para jovens é uma tendência mundial (CSTA, 2016) e proposto no Brasil pela BNCC (MEC, 2018). Atualmente, a computação também é uma competência fundamental para o mercado de trabalho, tornando essencial a integração do seu ensino na Educação Básica (EC, 2017). A Sociedade Brasileira de Computação (SBC) afirma que a computação deve ser ensinada desde o Ensino Fundamental como as outras ciências, para que a sociedade tenha cidadãos qualificados e preparados aos desafios do Século XXI (SBC, 2019). Com atividades que façam sentido para os alunos, busca-se popularizar competências de computação de forma ampla na população, instigando assim uma participação e inclusão maior de alunos nos cursos superiores da área de computação com oportunidades crescentes para a carreira profissional.

Na área de computação, atualmente o *Machine Learning* (ML), ou aprendizagem de máquina, está em destaque, sendo um dos principais responsáveis pelo crescimento e aproximação entre os campos de estatística e computação, necessitando a gerência de dados (AMAZON, 2019). Observa-se, além disso, que os conhecimentos de probabilidade e estatística para análise de dados ajudam a criar e validar modelos de *Machine Learning* capazes de fazer avaliações com maior precisão e poder estatístico.

À medida que os serviços de ML se tornam mais onipresentes, é importante esclarecer os conceitos básicos de ML para pessoas de todas as idades (HITRON et al., 2019). Inclusive, essas aplicações utilizando técnicas de *Machine Learning* estão inseridas no nosso dia a dia, como reconhecimento facial, mecanismos de pesquisa, sistemas de detecção de fraudes de cartão de crédito, *chatbots* e assistentes digitais, serviço de tradução entre outros, existem diversas técnicas de *Machine Learning*, sendo umas das principais a classificação de imagens.

Ensinar conceitos e técnicas fundamentais de Inteligência Artificial (IA), incluindo *Machine Learning* tem sido tradicionalmente feito apenas no ensino superior (TORREY, 2012) (MCGOVERN et al., 2011). Neste contexto, a iniciativa *The Artificial Intelligence (AI) for K-12 initiative* (AI4K12) (TOURETZKY

et al., 2019a) começou a desenvolver diretrizes para o ensino de IA na Educação Básica. As diretrizes são baseadas em um conjunto de grandes ideias, incluindo ensinar computadores a aprender com os dados, os desafios envolvidos em fazer com que os agentes de IA interajam naturalmente com os seres humanos e os efeitos positivos e negativos da IA na sociedade. Novos cursos de IA, ferramentas e tutoriais estão sendo lançados para ensinar IA em escolas, nos EUA, China, Reino Unido e em outros lugares, especialmente durante os últimos anos (MARQUES et al., 2020). Como o ensino de competências de *Machine Learning* ainda não é incluído na Educação Básica, a maioria das Unidades Instrucionais é proposta como atividades extracurriculares, oficinas, cursos, palestras, acampamentos de verão, desafios ou atividades individuais, vídeos e demonstrações (MARQUES et al., 2020). Várias Unidades Instrucionais concentram-se exclusivamente em níveis mais baixos de aprendizagem (lembança e compreensão), enquanto algumas Unidades Instrucionais também se aproximam do nível de aplicação, levando os alunos a criar seu próprio ML, modelo que visa o desenvolvimento de uma solução de ML para um problema na comunidade (DESAFIO FAMÍLIA IA, 2019). Nesse contexto, tipicamente se adotam projetos abordando partes e/ou todo o processo de ML incluindo a preparação de dados, treinamento de modelo de ML, e a avaliação do desempenho do modelo treinado. Alguns cursos levam o aluno também a refletir sobre o uso da IA na sociedade atual (ELEMENTOS DA AI, 2019). Outros abordam questões morais e o impacto da IA nos seres humanos (AIINSCHOOLS, 2019) (APPS FOR GOOD, 2019a) (READYAI, 2019) (TOURETZKY et al., 2019c).

Quanto às Unidades Instrucionais, observa-se atualmente escassez de informações sobre a avaliação da aprendizagem dos alunos, que é importante como *feedback* para o aluno e o instrutor, a fim de orientar o processo de aprendizagem (MARQUES et al., 2020).

1.2 PROBLEMATIZAÇÃO

Mesmo existindo algumas atividades extracurriculares no ensino de conceitos de ML nas Unidades Instrucionais, quase não há modelos comumente adotados de como avaliar a aprendizagem de competências de *Machine Learning* na Educação Básica (MARQUES et al, 2020). Poucos, como o Desafio Família de IA (2019) e as AIINSCHOOLS (2019) propõem uma rubrica para uma avaliação baseada em desempenho para analisar artefatos criados pelos alunos. Sakulkueakulsuk et al. (2018) alocam pontuações com base na precisão dos modelos ML desenvolvidos. Como alternativa, o Desafio Família AI (2019) e Elementos da IA (2019) também adotam *quizzes* ou exercícios para avaliação dos

alunos. Porém, observa-se que para a maioria dos modelos não é apresentado como foi sistematicamente definido e/ou avaliado, o que torna a confiabilidade e validade desses modelos de avaliação questionável.

Consequências dessa falta de modelos de avaliação também impactam no processo de aprendizagem do aluno, não fornecendo *feedback* instrucional formativo e somativo para guiar o seu desenvolvimento. O *feedback* deve assumir um papel formativo no contexto educacional, onde o principal objetivo é aumentar o conhecimento, habilidades e compreensão do aluno. Além de fornecer aos professores evidências sobre como os alunos estão aprendendo, para que eles possam revisar as instruções de forma apropriada, a avaliação formativa pode envolver diretamente os alunos no processo de aprendizagem, como por exemplo, fornecendo *feedback* que ajudará os alunos a entender como melhorar (ou implementar) ajustes instrucionais com base nos resultados da avaliação (BLACK & WILIAM, 1998a, 1998b; SHUTE, 2007).

Nesse âmbito, o problema é que atualmente não existem modelos de avaliação de aprendizagem de competências de *Machine Learning* dos alunos no contexto dos anos finais da Educação Fundamental e do Ensino Médio, com evidência de confiabilidade e validade.

Assim a pergunta de pesquisa do presente trabalho é: Como avaliar a aprendizagem de competências de *Machine Learning* dos alunos dos Anos Finais do Ensino Fundamental e do Ensino Médio de forma confiável e válida?

1.3 OBJETIVOS

Objetivo Geral

A pesquisa visa desenvolver um modelo de avaliação da aprendizagem de competências de *Machine Learning* nos Anos Finais do Ensino Fundamental e no Ensino Médio. O desenvolvimento inclui tanto a revisão sistemática de literatura quanto a definição sistemática de um modelo de avaliação, principalmente na avaliação com base no desempenho do aluno em Unidades Instrucionais voltadas ao ensino de ML para a classificação de imagens. Visa-se aplicar e avaliar o modelo de avaliação criado em relação a confiabilidade e validade convergente por meio de estudos de caso.

O trabalho é realizado no contexto da Iniciativa Computação na Escola/INCoD/INE/UFSC dando continuidade ao desenvolvimento de modelos de avaliação (ALVES et al., 2019) com enfoque na avaliação de aprendizagem de *Machine Learning*.

Objetivos específicos

Os objetivos específicos do presente trabalho são:

Objetivo I: Levantar o estado da arte em relação a modelos da avaliação de competências de ML na Educação Básica.

Objetivo II: Desenvolver um modelo de avaliação de competências de ML na Educação Básica decompondo sistematicamente a qualidade em características e definindo os indicadores e critérios de avaliação.

Objetivo III: Aplicar e Avaliar o modelo de avaliação desenvolvido em um estudo de caso em termos de confiabilidade e validade convergente.

1.4 METODOLOGIA DE PESQUISA

A pesquisa utilizada neste trabalho é caracterizada como uma pesquisa exploratória, com o objetivo de aprimorar ideias ou a descoberta de novas hipóteses a partir da análise do estudo da arte e de estudos de caso, possui natureza aplicada, haja vista compreende o desenvolvimento de um modelo conceitual de *Machine Learning* na Educação Básica. Adotando uma abordagem multimétodo, são utilizadas técnicas qualitativas e quantitativas.

Quadro 1 - Etapas da pesquisa

Etapa	Atividades	Métodos	Resultados
Etapa 1 Síntese da fundamentação teórica	Sintetizar o ensino de <i>Machine Learning</i> na Educação Básica	Pesquisa Bibliográfica (GIL, 2010)	Fundamentação teórica

Etapa 2 Levantamento do estado da arte	Levantar modelos de avaliação de aprendizagem de <i>Machine Learning</i> no contexto educacional	Mapeamento Sistemático da Literatura (PETERSEN et al. 2015)	Análise do estado da arte
Etapa 3 Desenvolvimento do modelo	Indicar o propósito da avaliação e os objetivos	Modelo ADDIE (BRANCH, 2009)	Modelo de avaliação desenvolvido
	Desenvolver critérios de avaliação para cada objetivo	Rubricas MOSKAL e LEYDENS (2000)	
	Revisar os critérios	Comparação com literatura	
	Instanciar o Modelo por meio de rubrica		
Etapa 4 Aplicação e avaliação do modelo de avaliação	Definir a avaliação	Estudo de caso (YIN, 2001) GQM (BASILI et al., 1994)	Modelo avaliado preliminarmente
	Aplicar o curso de <i>Machine Learning</i> para Todos da CnE e coletar dados		
	Analisar a validade e confiabilidade do modelo	- Coeficiente Ômega (FLORA, 2020) - Correlação Policórica (OLSSON, 1979)	

Fonte: elaborada pela autora

Síntese da fundamentação teórica: O objetivo desta etapa é a definição de conceitos de *Machine Learning* na Educação Básica. Para isso foi realizada uma pesquisa bibliográfica (GIL, 2010) e uma síntese teórica sobre conceitos relacionados ao escopo deste trabalho.

Levantamento do estado da arte: Nesta etapa é realizado um mapeamento sistemático que consiste em categorizar uma grande quantidade de estudos existentes na literatura (Petersen et al. 2015) com o objetivo de levantar o estado da arte sobre modelos existentes de análise automatizada de atividades de *Machine Learning* no contexto da Educação Básica. Na etapa de definição é estabelecida uma pergunta de pesquisa, seus objetivos e os critérios para inclusão e exclusão de resultados. É definida também a *string* de busca e as bases de dados nas quais a pesquisa é realizada. Como resultado, é definido o protocolo de busca. Na execução são realizadas buscas nas bases de dados a partir da *string* definida. Usando os critérios de inclusão/exclusão, trabalhos relevantes são selecionados. Na etapa de análise dos resultados, são extraídas informações dos trabalhos selecionados e é realizada uma análise e discussão identificando pontos fortes e fracos de cada trabalho.

Desenvolvimento do modelo: Seguindo o modelo de *design* instrucional ADDIE (BRANCH, 2009) e a abordagem de desenvolvimento de rubricas conforme proposto por MOSKAL e LEYDENS (2000) o modelo desenvolvido *Evidence centered design*.

Com base no *framework* CSTA (CSTA, 2016) e AI4k12 (TOURETZKY et al., 2019) são definidos os objetivos de aprendizagem a serem avaliados. A partir dessa definição são desenvolvidos critérios de avaliação para cada objetivo. O modelo é revisado em relação à literatura existente. A partir do modelo desenvolvido é instanciada uma rubrica levando-se em consideração também os resultados do mapeamento sistemático da literatura.

Aplicação e Avaliação do modelo de avaliação: O modelo desenvolvido é avaliado por uma série de estudos de caso (YIN, 2001). Utilizando o método *Goal Question Metric* (GQM) (BASILI; CALDIERA; ROMBACH, 1994) é definido o que deve ser avaliado em relação ao modelo desenvolvido e quais instrumentos de coleta de dados são utilizados. Para avaliar o modelo é aplicado o curso de *Machine Learning* da Computação na Escola (CnE) em instituições de ensino e em paralelo são coletados os dados. A partir desses dados é realizada uma análise estatística preliminar sobre a confiabilidade do modelo por meio do coeficiente Ômega (FLORA, 2020). Além disso, a validade convergente é analisada por meio da Correlação Policórica (OLSSON, 1979), e os dados são discutidos e interpretados de forma a identificar a confiabilidade e validade convergente do modelo desenvolvido.

1.5 CONTRIBUIÇÕES

Contribuição científica. A principal contribuição científica é um modelo de avaliação de aprendizagem de competências de ML no contexto da Educação Básica. Devido à falta de abordagens sistemáticas para a avaliação de aprendizagem, esse modelo representará uma contribuição inédita nessa área. Foi criado também um corpo de conhecimento em relação à confiabilidade e validade do modelo.

Contribuição social. Em relação ao âmbito social, espera-se que o trabalho possa despertar o interesse dos alunos em aprender os conceitos de Inteligência Artificial, especificamente os conceitos de *Machine Learning* desde a Educação Básica, além de contribuir na aprendizagem, consequentemente na qualidade do ensino, preparando-os para os desafios do século XXI.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 MACHINE LEARNING

De forma conceitual, compreende-se que *Machine Learning* (ML) permite que sistemas aprendam e melhorem automaticamente com experiências, sem serem programados explicitamente por meio do treinamento de um modelo a partir de dados que generalizam uma decisão contra uma medida de desempenho (MITCHELL, 1997). Algoritmos ML constroem um modelo matemático baseado em dados de amostra, conhecido sem "dados de treinamento", a fim de fazer previsões ou decisões sem serem explicitamente programados para executar a tarefa.

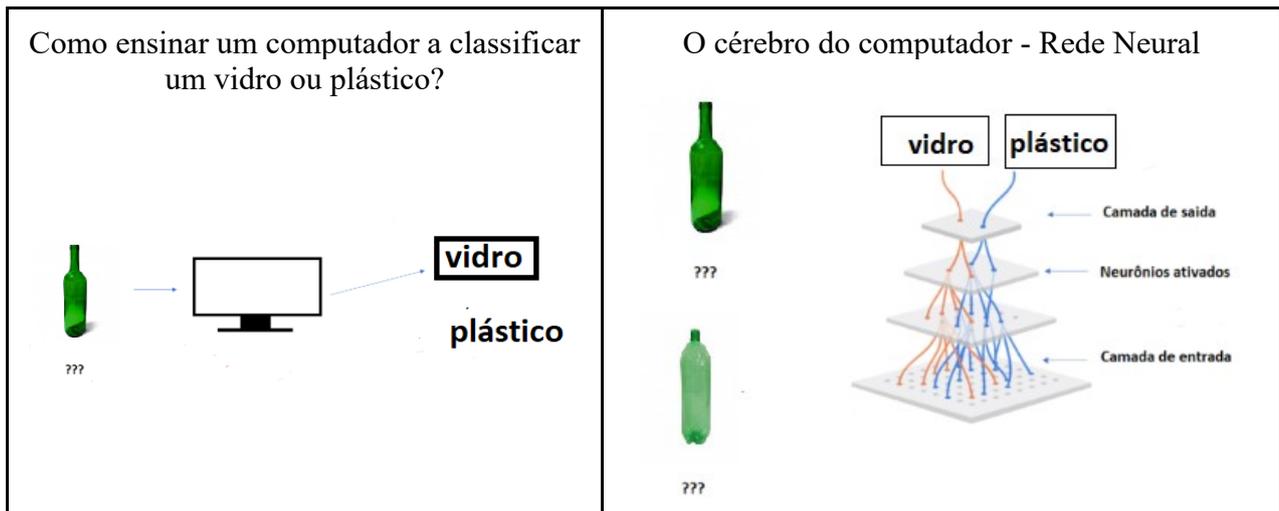
O *Machine Learning* é um subcampo da Inteligência Artificial (IA) na área de computação que permite que os computadores façam previsões e decisões para resolver problemas (AI4ALL,2020). Os modelos de *Machine Learning* (ML) são integrados em muitos sistemas e serviços que se tornam parte da vida cotidiana. Por exemplo, ao marcar fotos nas mídias sociais, o ML é usado para identificar rostos e, ao interagir com os serviços de assistente pessoal baseados em fala, o ML é usado para detecção de fala.

Redes Neurais

As redes neurais são formas de representação de conhecimento no *Machine Learning*, pois são grupos de algoritmos, vagamente modelados com base no cérebro humano, utilizados para reconhecimento de padrões. Elas interpretam dados sensoriais por meio de uma espécie de percepção de máquina, categorizando ou agrupando entradas puras de dados. Estes padrões reconhecidos por elas são numéricos, contidos em vetores, nos quais todos os dados do mundo real, sejam eles imagens, sons ou textos, precisam ser “traduzidos” (SKYMIND, 2019a).

Uma rede neural tenta imitar a maneira como um cérebro humano aborda problemas e usa camadas de unidades interconectadas para aprender e inferir relacionamentos com base nos dados observados, conforme o quadro 2.

Quadro 2 - Como funciona o cérebro do computador - Rede Neural



Fonte: elaborada pela autora

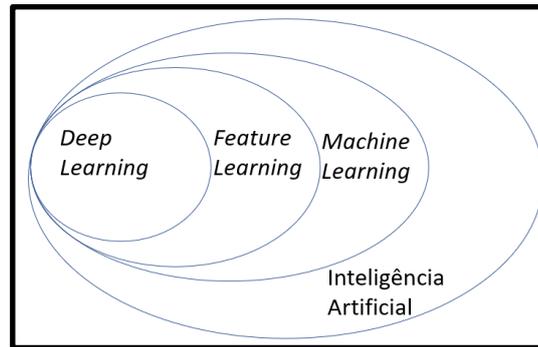
Uma rede neural pode ter várias camadas conectadas. Quando houver mais de uma camada oculta em uma rede neural, é chamada aprendizagem profunda e os modelos de redes neurais são capazes de ajustar e aprender conforme os dados mudam (HURWITZ e KIRSCH, 2018).

Os algoritmos para treinar redes neurais podem ser classificados em várias categorias de acordo com o estilo de aprendizagem (GOODFELLOW et al., 2016), a saber: Aprendizado supervisionado, semi-supervisionado, e não supervisionado. No aprendizado supervisionado, o algoritmo constrói um modelo matemático a partir de um conjunto de dados que contém tanto as entradas quanto as saídas desejadas. Algoritmos de classificação e algoritmos de regressão são tipos de aprendizado supervisionado. No aprendizado semi-supervisionado, uma combinação de dados rotulados e dados não rotulados é usada para fazer melhores previsões para novos pontos de dados do que usando apenas os dados rotulados. Algoritmos de aprendizagem não supervisionados são usados para encontrar estrutura/padrões nos dados, como agrupar os pontos de dados em categorias. Algoritmos de aprendizagem reforçado recebem *feedback* na forma de reforço positivo ou negativo em um ambiente dinâmico e são usados, por exemplo, em veículos autônomos.

***Deep learning*/Aprendizado profundo**

Deep learning é uma técnica que permite sistemas de computadores se aprimorarem por meio de experiências e dados (GOODFELLOW et al., 2016). *Deep learning* usa “redes neurais empilhadas”, ou seja, uma rede com múltiplas camadas de redes.

Figura 1 - Relação entre Deep Learning e a área de inteligência artificial



Fonte: adaptado a partir de GOODFELLOW et al., 2016

O aprendizado profundo é projetado para emular como o cérebro humano funciona para que os computadores possam ser treinados em lidar com abstrações e problemas mal definidos. Redes neurais e aprendizado profundo são frequentemente usados em reconhecimento de imagem, fala e aplicativos de visão computacional (HURWITZ e KIRSCH, 2018).

Nesse contexto, as técnicas do *Deep Learning* são tipos específicos de aprendizagem de representação muitos usados para fins de classificação (GOODFELLOW et al., 2016):

- Detecção de faces, identificação de pessoas em imagens, reconhecimento de imagens e expressões faciais (raiva, felicidade...);
- Identificação de objetos em imagens (sinais de rua, pedestres...);
- Reconhecimento de gestos em vídeos;
- Reconhecimento de voz, transcrever fala para texto, reconhecimento de sentimentos através da voz;
- Classificar textos como spam (em e-mails), ou fraudulentos (em reivindicações de seguros), reconhecimento de sentimentos através da escrita (*feedback* de professor e aluno).

No tocante ao tema reciclagem, um exemplo de uso de *Deep Learning* para fins de classificação é visto no estudo de Yang e Thung (2016), em que foi desenvolvido um sistema de classificação de lixo. Ao receber uma imagem de algum objeto o sistema classifica a categoria de lixo à qual ele pertence (papel, vidro, plástico ou metal).

Figura 2 - Sistema de *Deep Learning* capaz de identificar o tipo de lixo a partir de uma imagem



Fonte: YANG e THUNG, 2016

A criação de sistemas de *Deep Learning* é um processo iterativo que envolve uma sequência de etapas, que tipicamente incluem (AMAZON, 2019):

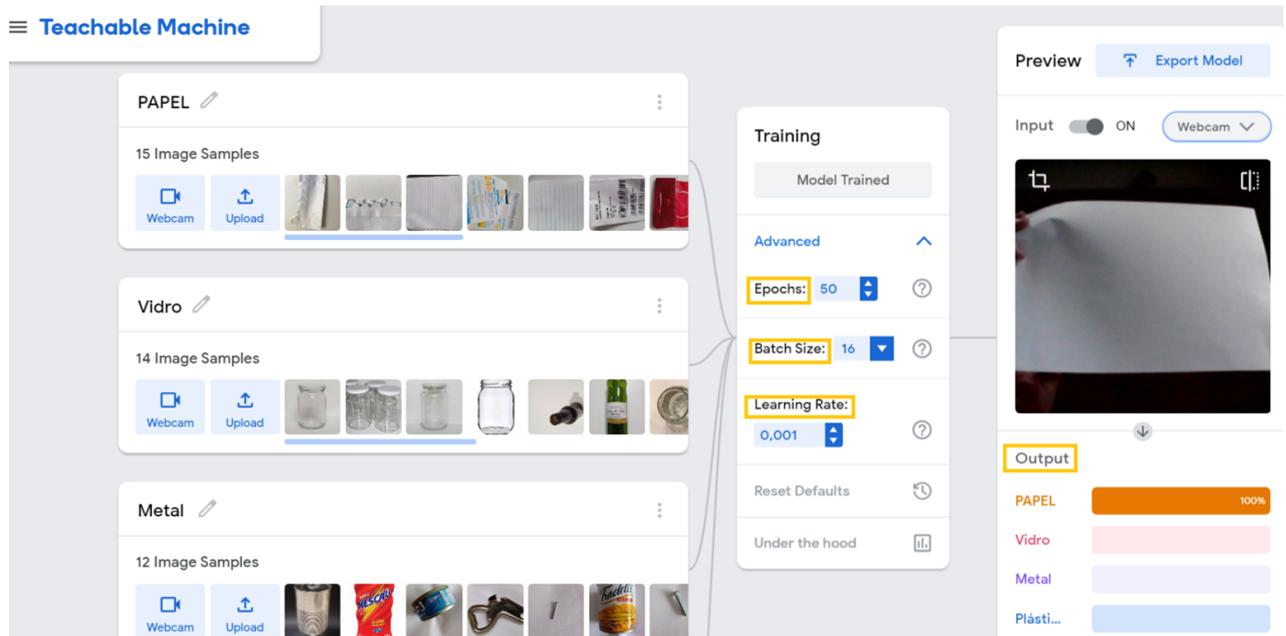
1. Análise de requisitos: Nesta etapa, o principal objetivo do modelo e seus recursos de destino são especificados. Isso também inclui a caracterização dos insumos e resultados esperados, especificando o problema.
2. Gerenciamento de dados: Durante a coleta de dados, os conjuntos de dados disponíveis são identificados e/ou dados são coletados. Isso também pode incluir a seleção de conjuntos de dados genéricos disponíveis, por exemplo, ImageNet, bem como conjuntos de dados especializados para transferência de aprendizado. O tipo de dado depende da tarefa de aprendizado de máquina, por exemplo, imagens, som, texto etc. Os dados são preparados validando e limpando os dados. Os conjuntos de dados podem ser rotulados em aprendizado supervisionado. O conjunto de dados geralmente é dividido em um conjunto de treinamento para treinar o modelo, conjunto de validação para selecionar o melhor candidato de todos os modelos e um conjunto de teste para realizar uma avaliação de desempenho imparcial do modelo escolhido em dados não vistos (RIPLEY, 2008).
3. Treinamento de modelo: Em seguida, um modelo é construído ou, mais tipicamente, escolhido dentre os conhecidos modelos comprovadamente eficazes em problemas ou domínios comparáveis, por exemplo,

(MODELZOO, 2019) alimentando os recursos/dados para o algoritmo de aprendizado durante o seu treinamento. O *Transfer Learning* potencialmente melhora os resultados de um modelo, mesmo com poucas imagens o modelo funciona bem e atende o objetivo, ou seja, atinge o resultado que se torna um indicador de sucesso na aprendizagem.

O treinamento de um modelo pode ser parametrizado tipicamente pela definição de:

- **Épocas** - Representa quantas vezes o conjunto de dados passa pela rede que está sendo treinada, para que os análise e aprenda. Cada vez que a rede vê uma entrada, como exemplo, de uma imagem, reconhece um pouco melhor. Porém quando se repete muitas vezes com um conjunto de dados pequenos pode ocorrer *overfitting* (é um termo usado para descrever quando um modelo estatístico se ajusta muito bem ao conjunto de dados anteriormente observado, mas se mostra ineficaz para prever novos resultados).
- **Batch Size** (Tamanho do lote) - É um subconjunto de dados. Como o conjunto de dados de treinamento pode ser grande, é dividido em partes (*batches*), para passarem pela rede de um em um. Um *epoch* está completo quando todos os *batches* passam pela rede uma vez.
- **Learning Rate** (Taxa de aprendizagem) - É o parâmetro utilizado para definir quão rápido a rede irá aprender. Uma taxa de aprendizagem baixa significa um aprendizado mais lento e com mais detalhes, mas o modelo levará muito tempo para aprender. E com a taxa de aprendizagem alta, apresenta rapidez no aprendizado, contudo com poucos detalhes.

Figura 3 - Parâmetros de treinamento do Google Teachable Machine



Fonte: Google Teachable Machine

Este exemplo usando a Google Teachable Machine (Figura 3) mostra a seleção de parâmetros de treinamento.

4. Avaliação do modelo: A qualidade do modelo treinado é avaliada para testar o modelo fornecendo uma melhor aproximação de como o modelo será executado no mundo real, por exemplo, analisando a correspondência entre os resultados do modelo e a rotulagem humana. A Figura 3, mostra que o modelo tem muita confiança de que o objeto na imagem é de papel. A qualidade dos modelos é avaliada para entender como melhorar seu desempenho, por exemplo, em termos de alta precisão, menor erro, testando o modelo contra dados não vistos anteriormente (THARWAT, 2019).

As medidas tipicamente usadas para esta avaliação, especialmente no contexto de modelos para classificação de imagens, incluem:

Acurácia: A taxa de acurácia é usada para avaliar o quanto a rede está acertando em suas classificações. Quando a taxa de acurácia é 1 indica que o modelo está perfeito, pois acertou a classificação para todas as imagens do conjunto de teste.

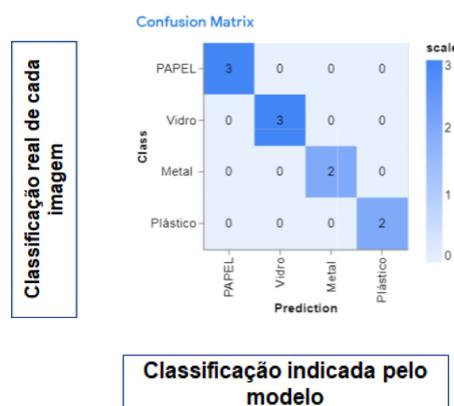
Figura 4 - Taxa de Acurácia

Accuracy per class		
CLASS	ACCURACY	# SAMPLES
PAPEL	1.00	3
Vidro	1.00	3
Metal	1.00	2
Plástico	1.00	2

Fonte: Google Teachable Machine

Matriz de confusão: A matriz de confusão é usada para avaliações de modelos de classificação em *Machine Learning*. É uma tabela que mostra as frequências de classificação para cada classe do modelo. O número de acertos, para cada classe, se localiza na diagonal principal da matriz, os demais elementos quando maiores que zero, fora da diagonal principal, representam erros na classificação. A matriz de confusão de um modelo ideal possui números maiores que zero apenas na diagonal principal, e fora da diagonal todos os elementos iguais a zero, uma vez que ele não comete erros. O exemplo apresentado na Figura 5, nos mostra que as frequências de classificação estão perfeitas, nenhuma imagem foi classificada erroneamente em outra categoria.

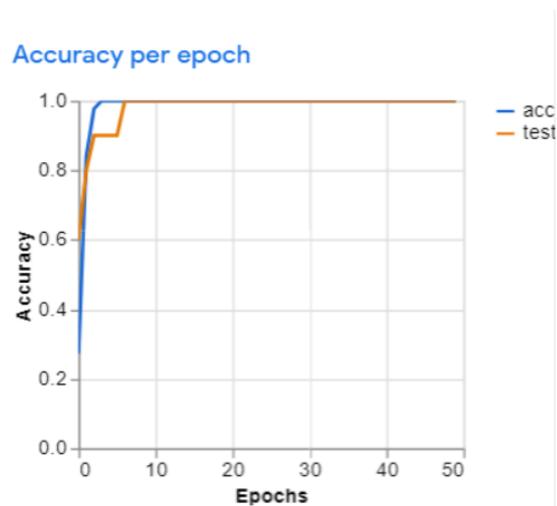
Figura 5 - Matriz de confusão



Fonte: Google Teachable Machine

- **Acurácia durante o treinamento:** A Figura 6 mostra a precisão das classes ao longo das 50 temporadas, não tendo alterações após a 5ª época. Ou seja, poderíamos ter executado somente 6 temporadas para chegar no mesmo resultado.

Figura 6 - Precisão por época durante o treinamento



Fonte: Google Teachable Machine

- *Loss Function*: A função de perda é usada para determinar o erro entre a saída dos algoritmos e o valor alvo fornecido.

5. Implantação do modelo: Durante a fase de produção, o modelo é integrado em um produto de *software*, como por exemplo sistema web ou aplicativo móvel, com intuito de aplicá-lo a novos dados recebidos em tempo real.

Existem diversas ferramentas que suportam o processo de ML, inclusive ferramentas visuais que não necessitam de codificação (GRESSE VON WANGENHEIM et al., 2021). Um exemplo é o Google Teachable Machine (TEACHABLE MACHINE, 2021), uma ferramenta baseada na web que torna a criação de modelos de aprendizado de máquina rápida, fácil e acessível a todos os alunos. Pode usar arquivos ou capturar exemplos ao vivo, também pode optar por usá-lo inteiramente no dispositivo, sem quaisquer dados de webcam ou microfone do computador. O GTM permite a classificação de:

- Imagens: Criar um modelo a classificar imagens usando arquivos ou webcam;
- Sons: Criar um modelo para classificar áudio gravando amostras curtas de som. (Suporte a arquivos WAV/MP3/etc);

- Poses: Criar um modelo para classificar as posições do corpo usando arquivos ou poses marcantes em sua webcam.

O modelo pode ser treinado e testado instantaneamente, agindo conforme o esperado, pode ser aplicado em serviços como sites, sistemas, aplicativos etc., também pode baixar o modelo ou hospedá-lo online gratuitamente. O sistema utilizado, Google Teachable Machine, possibilita que os estudantes avaliem a qualidade do modelo treinado em relação a acurácia, matriz de confusão e com base na interpretação destas avaliações, realizem ajustes no treinamento do modelo de ML caso necessário.

2.2 ENSINO DE *MACHINE LEARNING* NA EDUCAÇÃO BÁSICA

Considerando a importância do *Machine learning* (ML) no dia a dia, surgem esforços para incluir o ensino de ML na Educação Básica. Atualmente um dos principais esforços é coordenado pelo Grupo de Trabalho K-12 (AI4K12), iniciativa conjunta da Associação para o Avanço da Inteligência Artificial (AAIA) e da Associação dos Professores de Ciência da Computação (CSTA) que visa desenvolver diretrizes do ensino de inteligência artificial para os alunos da Educação Básica.

Os princípios para refinamento e escopo das diretrizes precisam ter relevância no mundo real para permitir que os alunos possam adquirir novas competências que serão necessárias para um futuro em que a inteligência artificial transforma a maneira como nos comunicamos, trabalhamos e vivemos uns com os outros e com as máquinas. Definida a alfabetização da IA como um conjunto de competências que permite ao indivíduo avaliar criticamente as tecnologias de IA; comunicar e colaborar efetivamente com a IA; e usar IA como uma ferramenta (LONG e MAGERKO, 2020).

Quadro 3 - Competências para avaliar as Tecnologias de IA

Competência		Descrição
1	Reconhecendo IA	Distinguir entre artefatos tecnológicos que usam ou não IA
2	Entendendo Inteligência	Analisar e discutir criticamente características que fazem uma entidade “inteligente”, incluindo discutir diferenças entre humanos, animais e inteligência de máquina

Competência	Descrição
Competência 1 (reconhecimento da IA)	É um fator crítico a habilidade necessária para interações com a IA.
Competência 2 (compreensão da inteligência)	Atividades como comparar dispositivos de IA e IA vs. habilidades humanas foram usadas para promover esse entendimento. Tomadas em conjunto com pedidos recentes de IA ampliada nos currículos.
Competência 3 (interdisciplinaridade)	Sugerem a importância de entender que a IA é interdisciplinar.
Competência 4 (geral x estreita)	Das três áreas da IA - Sistemas cognitivos, robótica e máquina aprendizagem, produziram "IA estreita", ou AI que seja inteligente em um domínio específico, mas "IA geral", ou IA que rivaliza com a inteligência humana, domínios múltiplos, ainda não foi alcançado.
Competência 5 (pontos fortes e fracos da IA)	Além das cinco "grandes ideias" da IA que podem ajudar os usuários a tomar decisões mais informadas, os estudantes provavelmente estarão mais bem equipados para alavancar as diferentes capacidades da IA e dos seres humanos para resolver problemas se eles entenderem os pontos fortes e fracos da IA.
Competência 6 (imaginar "IA futura")	A capacidade de imaginar "IA futura" pode permitir que indivíduos explorem criativamente novas ideias, considerem os valores inerentes a uma tecnologia e avaliem criticamente os efeitos a longo prazo que uma tecnologia pode ter no mundo.
Competência 7 (representações)	O conceito de compreensão das representações, uma das “grandes ideias” de AI, poderia ajudar os usuários a entender como os computadores representam o conhecimento e reconhecem que algum conhecimento é sempre perdido em uma representação do mundo.
Competência 8 (tomada de decisão)	Os sistemas cognitivos usam muitas estratégias para planejamento, decisão, criação, resolução de problemas e aprendizado. Provavelmente não é necessário o usuário entender todas essas estratégias em detalhes, mas um entendimento de alto nível de como os computadores tomam decisões pode auxiliar na interpretação e compreensão de algoritmos.
Competência 9 (etapas do ML)	Alguns trabalhos recentes estão começando a investigar como ensinar ML para indivíduos sem formação em Ciências da Computação. Algumas dessas iniciativas se concentram no ensino não especialistas em como implementar algoritmos ML; outros se concentram na comunicação de práticas de mais alto nível, como dados coleta e preparação, seleção de modelos, treinamento, teste e previsão.
Competência 10 (recursos)	Os alunos também costumam surpreender que o ML exige tomada de decisão humana e não é totalmente automatizado. Pesquisas também sugerem que uma

humanos - Papel na IA)	maneira de dissipar os alunos dos conceitos errôneos sobre ML é envolver-se em interação, incluindo projetos em que os alunos possam treinar modelos de ML.
Competência 11 (alfabetização de dados)	Definem um conjunto de competências para alfabetização de dados (por exemplo, a “capacidade de avaliar criticamente dados e suas fontes”). A importância dessas competências para entender ML sugere que o conhecimento dos dados básicos, conceitos científicos é um componente da alfabetização em IA (Prado e Marzal, 2013).
Competência 12 (aprendendo com os dados)	Reconhecendo quando os dados pessoais estão sendo usados para treinar ML e interpretar os resultados de algoritmos no contexto dos dados em que foram treinados são duas questões de alfabetização de dados particularmente relevantes para a IA. Pesquisa sugere que é importante que os alunos entendam que computadores aprendem com seus dados.
Competência 13 (interpretação crítica de dados)	Os alunos devem examinar criticamente os dados com “ceticismo e interpretação”.
Competência 14 (ação / reação)	Conceitos como controle reativo e compreensão da percepção e ação de sensores são específicos para robótica. Entender que os agentes de IA podem agir fisicamente e reagir ao mundo é um importante pré-requisito para entender a robótica.
Competência 15 (sensores)	Aprendendo sobre sensores e suas capacidades (uma das “grandes ideias” da IA) também podem ajudar a entender como os dispositivos de IA reúnem dados e interface com o mundo.
Competência 16 (ética)	Nas diretrizes atuais para os currículos de graduação em ciências da computação é importante incluir um curso de ética em que os alunos aprendam sobre ética e como aplicá-la para avaliar a tecnologia. Essas habilidades podem ajudar tanto os profissionais de computação quanto os usuários comuns na identificação de quando é apropriado usar IA.
Competência 17 (programabilidade)	Pesquisas indicam que as crianças primeiro personificam agentes e então reconhecem que eles são programáveis. Este reconhecimento é fundamental para entender como a IA funciona. Fornece oportunidades e promove esse entendimento para alunos de todas as idades programarem a IA.

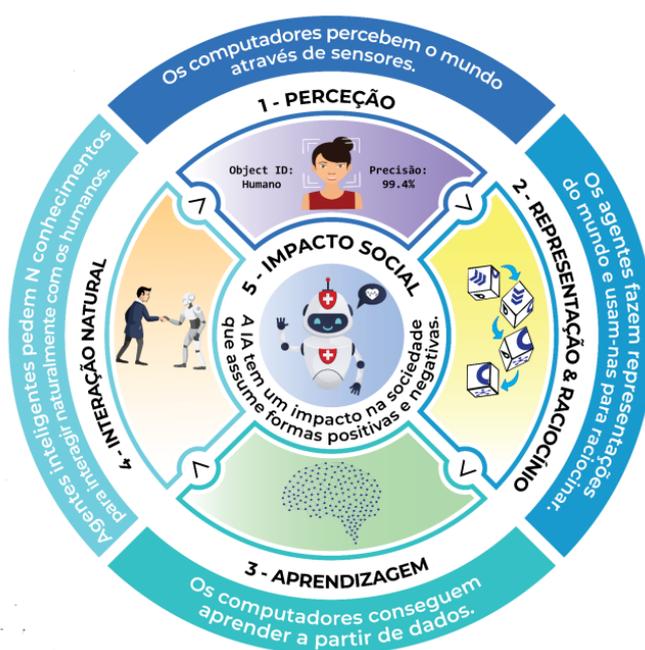
Fonte: Long e Magerko, 2020

Existem várias razões para iniciar o processo de ensino de Inteligência Artificial/*Machine Learning*, na Educação Básica, que comprovam os aspectos que já existem e outros que em breve serão questões ativas na sociedade (TOURETZKY et al., 2019):

- Explicar como funciona um carro autônomo e os tipos de subsistemas de IA envolvidos em seu funcionamento.
- Explicar o processo pelo qual os modelos de ML são desenvolvidos, a partir da coleta de dados para treinamento e teste, fontes etc.
- Use, modifique e crie sistemas de IA usando ferramentas apropriadas para o desenvolvimento.
- Compreender as implicações da IA para problemas do mundo real.

De acordo com AI4K12 [Touretzky et al. 2019c], cinco grandes ideias devem ser consideradas neste estágio escolar: percepção, representação e raciocínio, aprendizagem, interação natural e impactos sociais, incluindo na aprendizagem a *Machine Learning* (ML) (WOLLOWSKI et al. 2016) (TOURETZKY et al. 2019a) (Figura 7).

Figura 7 - Cinco grandes ideias em Inteligência Artificial



Fonte: Touretzky, 2019

Essas cinco grandes ideias também incluem o ensino sobre o *learning*, que se refere ao conteúdo de *Machine Learning*. De acordo com AI4K12, os conceitos de *Machine Learning* a serem abordados no ensino básico devem incluir (TOURETZKY et al., 2019c):

O que é aprendido?

- Tipos de algoritmos de aprendizado por estilo
- Fundamentos das redes neurais
- Tipos de arquitetura de rede neural
- Como o treinamento de dados influencia o aprendizado
- Limitações do aprendizado de máquina

Entre esses conceitos de *Machine Learning*, cada um é projetado para uma determinada tarefa e objetivo.

Quadro 4 - Rascunho do Conteúdo da grande ideia 3 - Progressão

Grande Ideia #3	OA = Objetivo de Aprendizagem: O que os alunos devem ser capazes de fazer.	
Conteúdos / Conceito	3-5 (anos iniciais do fundamental)	6-8 (anos finais do fundamental)
Natureza da Aprendizagem (Humanos vs. máquinas) 3-A-i	OA: Diferencie como as pessoas aprendem e como os computadores aprendem.	OA: Compare as características únicas do ser humano para aprender com os sistemas de aprendizado de máquina como operar.
Natureza da Aprendizagem (Encontrando padrões em dados) 3-A-ii	OA: Modelar como a aprendizagem supervisionada identifica padrões em dados rotulados.	OA: Modelo de como a aprendizagem não supervisionada encontra padrões em dados não rotulados.
Natureza da Aprendizagem (Treinando um modelo) 3-A-iii	OA: Treine um modelo de classificação usando aprendizado de máquina, e, em seguida, examine a precisão do modelo em novas entradas.	OA: Treine e avalie uma classificação ou previsão modelo usando aprendizado de máquina para tabular. um conjunto de dados
Natureza da Aprendizagem (Construindo vs. usando um raciocinador) 3-A-iv	OA: Demonstrar como os dados de treinamento são rotulados quando se usa uma ferramenta de aprendizado de máquina.	OA: Explique a diferença entre treinar e usar um modelo de raciocínio.
Natureza da Aprendizagem (Ajustando representações interna) 3-A-v	OA: Análise um jogo onde se constrói uma árvore de decisão, descrevendo a organização da árvore e o algoritmo de aprendizagem usado para adicionar nós.	OA: Compare como um algoritmo de aprendizagem de árvore de decisão funciona vs. como um algoritmo de aprendizagem de rede neural trabalha.
Natureza da Aprendizagem (Aprendendo com experiência) 3-A-vi	OA: Explique como a aprendizagem por reforço permite um computador aprender com a experiência (ou seja, tentativa e erro).	OA: Explique as diferenças entre a aprendizagem supervisionada e aprendizagem por reforço.
Redes neurais (Estrutura de um sistema rede neural) 3-B-i	OA: Ilustre como uma rede neural de 1 a 3 neurônios é uma função que calcula uma saída.	OA: Ilustrar a estrutura de uma rede neural e descrever como suas partes formam um conjunto de funções que calculam uma saída.
Redes neurais (Ajuste de peso) 3-B-ii	OA: Demonstrar como os pesos são atribuídos em um sistema de rede neural para produzir um comportamento de entrada / saída desejado.	OA: Demonstrar como uma regra de aprendizagem pode ser usada para ajustar os pesos em uma rede neural de uma camada.

Conjuntos de dados (Conjuntos de recursos) 3-C-i	OA: Crie um conjunto de dados rotulado com recursos explícitos de vários tipos e use uma ferramenta de aprendizado de máquina para treinar um classificador nestes dados.	OA: Criar um conjunto de dados para treinar uma árvore de decisão classificador ou preditor e explorar o impacto que diferentes codificações de recursos têm na árvore de decisão.
Conjuntos de dados (Grandes conjuntos de dados) 3-C-ii	OA: Ilustrar como treinar um classificador para um conceito amplo como "cachorro" requer uma grande quantidade de dados para capturar a diversidade do domínio.	OA: Ilustrar como os objetos em uma imagem podem ser segmentados e rotulados para construir um conjunto de treinamento para o objeto de reconhecimento.
Conjuntos de dados (Viés) 3-C-iii	OA: Examinar recursos e rótulos de dados de treinamento para detectar fontes potenciais de viés.	OA: Explique como a escolha dos dados de treinamento molda o comportamento do classificador e como o viés pode ser introduzido se o conjunto de treinamento não estiver devidamente balanceado.

Fonte: AI4K12.org - versão traduzida

Para suportar a progressão da aprendizagem dos alunos nesta área de conhecimento complexa, o ciclo "*Use-Modify-Create*" (LYTLE et al., 2019; LEE et al., 2011) é proposto como base para a progressão dos conceitos e práticas de aprendizagem de *Machine Learning*. Após esse ciclo, os alunos se adaptam ao primeiro tópico "Use" e após analisar um determinado artefato, "Modify" um existente e, finalmente, "Create" novos.

Em todos esses níveis os alunos são levados a aprender e aplicar os conceitos de ML para desenvolver modelos de ML.

Para tornar o ensino de *Machine Learning* relevante aos alunos, a Unidade Instrucional segue os conceitos de Ação Computacional, um *framework* que propõe aliar o aprendizado de computação com o desenvolvimento de ações de impacto na sociedade e na vida do estudante (TISSENBAUM, SHELDON e ABELSON, 2019, p.34). Ao conectar com a vida real dos estudantes, pode-se ajudá-los a desenvolver uma consciência crítica do papel que eles podem exercer para afetar positivamente suas comunidades através da computação e os empoderar para fazer mais do que apenas aprender a programar. Logo, pode-se perguntar o que eles querem desenvolver e o motivo pelo qual querem desenvolver (TISSENBAUM, SHELDON e ABELSON, 2019, p. 34).

2.3 CURSO DE *MACHINE LEARNING* PARA TODOS!

Visando o ensino de *Machine Learning* na Educação Básica, foi desenvolvido pela iniciativa Computação na Escola/INCoD/INE/UFSC um curso de *Machine Learning* da Computação na Escola (MARQUES et al., 2020).

Contexto do curso

O curso de *Machine Learning* da Computação na Escola foi definido para o seguinte contexto:

Alunos: O público-alvo são alunos da escola pública dos Anos Finais do Ensino Fundamental e do Ensino Médio, com idade entre 12 e 18 anos. A maioria dos alunos nessa idade já possui conhecimentos prévios e habilidades no uso de computadores e outros dispositivos que acessam a internet. A maioria deles têm seu próprio telefone celular (predominantemente Android) e frequentemente passa muito tempo online, especialmente em redes sociais e/ou jogos digitais. Em relação às habilidades relacionadas à computação, especificamente a programação, alguns alunos já têm uma compreensão de programas de computador, mas poucos sabem como criá-los (FERREIRA et al., 2020).

Ambiente escolar: As aulas ocorrem tipicamente em Salas Informatizadas com computadores e acesso à internet. Na Rede Municipal de Florianópolis as Salas Informatizadas são utilizadas por professores de outras áreas não relacionadas a informática (português, matemática, ciências, geografia etc.), com o acompanhamento do Professor Auxiliar de Tecnologia Educacional para a realização de projetos educativos e trabalhos digitais ensinando alfabetização de TI, como edição de textos, apresentações, animação (Stop Motion), programação (CODE.ORG) etc. Tipicamente as salas informatizadas (SI) possuem de 12 a 20 computadores. As aulas ministradas nas SI possuem duração de 45 minutos, e são realizadas para turmas de aproximadamente 35 alunos.

A maioria dos Professores Auxiliares de Tecnologia Educacional das salas informatizadas não possuem formação na área de computação, geralmente possuem Pedagogia ou Licenciaturas com especializações em Tecnologias Educacionais, como pode-se observar o número baixo de formandos em cursos de licenciaturas de computação em comparação a outras disciplinas (Pelle, 2018).

Ambiente curricular. Com a intenção dos alunos compreenderem os fundamentos da Inteligência Artificial e Robótica, a BNCC apresenta proposta de inclusão no currículo a partir do ensino médio, não citando o ensino

de *Machine Learning* (BNCC II, 2021).

Currículo de *Machine Learning*. Espera-se que o ensino de *Machine Learning* seja inserido em uma Unidade Instrucional com o objetivo de avaliar os conceitos do aprendizado de *Machine Learning* em atividades conduzidas para alunos da Educação Básica. A unidade deve ser baseada nas principais diretrizes de currículo voltadas ao ensino de *Machine Learning* e design na Educação Básica (CSTA, 2016); AI4ALL (TOURETSKY et al., 2019); (SBC, 2018); (LONG e MAGERKO, 2020).

Objetivos de aprendizagem do curso.

O curso introduz conceitos básicos de *Machine Learning*, como algoritmos de aprendizagem e fundamentos de redes neurais, bem como discute as limitações e conceitos éticos relacionados a *Machine Learning*. O curso também ensina a aplicação desses conceitos, levando o estudante a criar o seu primeiro modelo de *Machine Learning* de reconhecimento de imagens para separação de lixo reciclável. Com relação à *Machine Learning*, são previstos objetivos de aprendizagem alinhados às diretrizes da educação básica *Guidelines for Artificial Intelligence* (TOURETZKY, 2019) referentes à *Big Idea 3 - Learning* e, também, alguns tópicos propostos pelo CSTA (CSTA, 2017).

Quadro 5. Objetivos de aprendizagem do curso

ID	Objetivo de aprendizagem	Área de conhecimento	Referência
OA1	Conhecer e identificar exemplos da aplicação de ML	ML	(AI4K12, 2019)
OA2	Descrever conceitos básicos de ML: o que é aprendizagem, fases do aprendizado computacional	ML	(AI4K12, 2019)
OA3	Identificar padrões em dados	ML	(AI4K12, 2019)
OA4	Simular o treinamento de uma rede neural simples	ML	(AI4K12, 2019)
OA5	Coletar e rotular dados para o treinamento de um modelo de ML pré-treinado; compreender como algoritmos de <i>Machine Learning</i> são influenciados pelos dados	ML	--Com base no processo de ML--
OA6	Modificar um projeto de <i>Machine Learning</i> interativo treinando o modelo	ML	(AI4K12, 2019)
OA7	Avaliar o desempenho do modelo de ML	ML	--Com base no processo de ML--

OA8	Discutir o impacto do <i>Machine Learning</i> na sociedade e o quanto seu desenvolvimento está ligado à ética	Impactos da computação	(CSTA, 2017: 1B-IC-18)
-----	---	------------------------	------------------------

Fonte: Computação na Escola/INCoD/INE/UFSC, Florianópolis, 2020.

Estratégia pedagógica

O curso pode ser aplicado em um modo de aprendizagem presencial/híbrido ou remoto (com ou sem instrutor), como um curso extracurricular ou em uma forma interdisciplinar como parte das aulas de ciências. Apresenta os conteúdos de *Machine Learning* na compreensão e utilização de um sistema de ML com uma abordagem que faz sentido no contexto dos alunos, tanto por ser realizado de forma interdisciplinar quanto por abordar um tema ainda problemático no Brasil e na comunidade local: a separação do lixo reciclável. De acordo com a BNCC, no ensino fundamental os alunos começam a compreender sua relação com o entorno, experimentando e estudando noções sobre a luz, som, calor e outros elementos, “além de prever a construção coletiva de propostas de reciclagem e reutilização de materiais” (MEC, 2018). Portanto, o curso exemplifica a aplicação de um modelo de ML por meio do Google Teachable Machine, que permite classificar resíduos/lixo em categorias de material reciclável (papel, vidro, plástico e metal). O presente curso aborda a primeira etapa (*Use*), podendo ser continuada por outra Unidade Instrucional em sequência, voltada à modificação e criação de um modelo de ML (*Create*). Com base na análise de contexto, a Unidade Instrucional é projetada como um curso *online* de 8 horas, conforme plano de ensino (Quadro 6). É adotada uma metodologia ativa de ensino de ML de forma interdisciplinar à disciplina de ciências, abordando o tema reciclagem na aplicação do curso. Também é previsto como atividade extra a implantação do modelo de ML criado com auxílio do Google Teachable Machine.

Programa do curso.

O curso está organizado nas seguintes partes:

- Noções gerais e relevância do ML
- Conceitos fundamentais
- Faça seu primeiro modelo de ML
- Revisão de conteúdo e processo de ML
- Questões éticas e impacto social do ML

Na primeira parte, AI / ML e sua aplicação na vida cotidiana são motivadas e ilustradas por meio de exemplos de demonstração prática. Os alunos são estimulados a discutir os aplicativos de IA presentes em seu cotidiano para perceber a relevância desse conhecimento. A segunda parte, sobre os conceitos fundamentais, apresenta os principais conceitos e ideias de ML e redes neurais. Isso também inclui a ilustração de gerenciamento de

dados e simulação de aprendizagem em um sistema neural simples, exemplo de rede. A parte “Faça seu primeiro modelo de ML!” orienta os alunos a desenvolver um modelo de ML para reconhecimento de imagem usando o Google Teachable Machine. Com base em um conjunto de dados pré-montados (com imagens "bagunçadas"), os alunos preparam o conjunto de dados, organizando / rotulando as imagens para cada categoria de reciclagem, bem como limpando o conjunto de dados. Em seguida, eles treinam o modelo. Depois de treinarem, eles testam o modelo com novas imagens carregadas ou capturadas via *webcam* e analisam as medidas de desempenho, incluindo a tabela de precisão e a matriz de confusão apresentada pelo Google Teachable Machine. Interpretando os resultados de desempenho, os alunos, caso necessário, ajustam o conjunto de dados e/ou modelo parâmetros de treinamento, a fim de melhorar o desempenho do modelo. Depois de aplicado o processo de ML e os conceitos básicos, as principais fases do processo são revisadas. A última parte do curso inclui uma discussão crítica e reflexão sobre questões éticas e sociais, o impacto do ML, bem como a motivação quanto às oportunidades de carreira nesta área.

Quadro 6 - Plano de ensino

Aula	Conteúdo	Objetivo de aprendizagem	Modo	Método instrucional	Material instrucional	Avaliação
Motivação						
1	Motivação sobre Inteligência Artificial e sua aplicação no dia a dia	OA1, OA8	Aula síncrona	Aula expositiva, discussão e demonstrações práticas	Slides 1 Demonstração prática: Object Detector and Classifier app, jogo Quick, draw! Material h5p	Quiz h5p
Conceitos fundamentais						
2	Conceitos básicos de ML e redes neurais	OA2, OA3, OA4	Aula assíncrona	Atividade interativa	Material h5p	Quiz h5p
Faça um modelo de ML						
3	Preparação de dados	OA3, OA5, OA9	Aula assíncrona acompanhamento no google meet sob demanda	Atividade interativa	Slides 3 Dataset Material h5p	Quiz h5p Avaliação de desempenho
4	Treinamento do modelo e avaliação	OA4, OA6, OA7	Aula síncrona	Atividade interativa	Slides 4 Material h5p Google Teachable Machine Questionário de avaliação de desempenho	Quiz h5p Avaliação de desempenho
5	Revisão do processo de ML	OA1, OA2, OA5	Aula síncrona	Discussão/dúvidas e motivação de preencher o questionário	Slides 5 Material h5p	Quiz h5p
Questões éticas e impacto						
6	Questões éticas referentes à IA/ML, limitações, perigos, oportunidades de emprego	OA1, OA8	Aula assíncrona	Atividade interativa	Slides 6 Demonstração prática: Moral Machine	Quiz h5p

					Material h5p	
7	Fechamento do curso:Discussão/Dúvidas	--	Aula síncrona	Discussão/dúvidas e motivação de preencher o questionário	Questionário de avaliação do curso	--

Fonte: Elaborado pela autora

Materiais didáticos. Os materiais disponibilizados para aplicação do curso incluem o plano de ensino, slides, materiais de apoio para as atividades práticas e de avaliação de aprendizagem. A avaliação de aprendizagem, por sua vez, se dá ao fim de cada aula por meio de perguntas objetivas, a fim de monitorar o avanço na aprendizagem (MARQUES, S. L., GRESSE VON WANGENHEIM, C, 2020) (<https://osf.io/preprints/socarxiv/wj5ne/>). O material didático do curso também inclui slides para as aulas síncronas como material interativo usando H5P (Figura 8). O H5P capacita todos os usuários a criarem experiências ricas e interativas de forma eficiente na web, também facilita a criação, compartilhamento, reutilização de conteúdo e aplicativos HTML5 (h5p.org). Para possibilitar o desenvolvimento do modelo de ML de forma fácil é adotada a Google Teachable Machine (<https://teachablemachine.withgoogle.com/>), uma ferramenta visual gratuitamente disponível online. Resultados da aprendizagem são avaliados por meio de *quizzes* interativos e são coletados artefatos criados pelos alunos via questionários online. O material interativo do curso foi disponibilizado via Moodle Grupos da Universidade Federal de Santa Catarina (UFSC) durante a aplicação do curso. O material interativo está também publicamente disponível no site dos cursos da iniciativa Computação na Escola (<https://cursos.computacaonaescola.ufsc.br/cursos/curso-mlparatodos>).

Figura 8- Exemplos do material didático

Fonte: Elaborada pela autora

Avaliação de aprendizagem. A avaliação da aprendizagem dos alunos é prevista de diferentes formas (Quadro 7).

Quadro 7 - Avaliação de aprendizagem no curso ML para Todos!

ID	Objetivo de aprendizagem	Área do conhecimento	Nível de aprendizagem do Bloom	Método de avaliação	Instrumento
OA1	Conhecer e identificar exemplos de aplicação do ML	ML	Compreensão	Teste simples (Quiz)	Quiz aula 1-4, 1-10, 1-20, 1-22
OA2	Descrever os conceitos básicos de ML: o que é uma rede neural, aprendizagem, processo ML	ML	Compreensão	Quiz	Quiz aula 2 - 25
OA3	Identificar padrões de dados	ML	Compreensão e aplicação	Drag - and - drop, rubrica	Quiz aula 3-10, 3-17, 3-26, 3-27, 3-28, 3-29, 3-30
OA4	Simular o treinamento de uma rede neural simples	ML	Aplicação		
OA5	Coletar, limpar e rotular dados para o treinamento de um	ML	Compreensão		Avaliação por desempenho

	modelo de ML; entender como os algoritmos de ML são influenciados pelos dados				
OA6	Treinar o modelo de ML	ML	Compreensão e aplicação	Drag - and - drop, múltipla escolha Avaliação por desempenho (parâmetros do treinamento)	Treinamento do modelo Quiz aula 4-11, 4-12, 4-13, 4-14 ,4-15, 4-16, 4-21, 4-22, 4-27, 4-28, 4-40 Avaliação por desempenho
OA7	Avaliar o desempenho do modelo de ML	ML	Compreensão e aplicação		Avaliação por desempenho
OA8	Discutir questões éticas e o impacto do ML na sociedade	Impactos da computação / IA	Conhecimento e compreensão		Aula 6 – Questões éticas e oportunidades

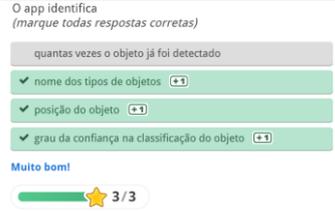
Fonte: Elaborado pela autora

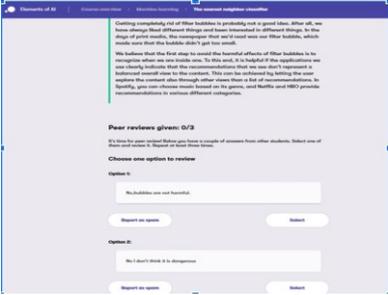
As atividades práticas de Unidades Instrucionais utilizam a abordagem de atividades abertas a partir de um tema predefinido. As atividades abertas possuem várias possíveis soluções, além disso o julgamento sobre qual caminho tomar para chegar a uma solução cabe ao aluno (LYE; KOH, 2014; GIJSELAERS, 1996). A vantagem de utilizar atividades abertas é que se incentiva e recompensa a criatividade e o pensamento analítico dos alunos (SURIF; IBRAHIM; DALIM, 2014). A avaliação em formato de *quiz* também faz parte da avaliação formativa, auxilia positivamente no processo de ensino e avaliação da aprendizagem dos estudantes, e são abordados aplicativos que podem ser usados para elaborar o *quiz*. O professor pode elaborar questões pertinentes ao conteúdo e socializar as questões com os grupos.

A avaliação por atividades online pode ser realizada por meio de diversas possibilidades que permite o acompanhamento das participações dos alunos nas atividades propostas pelo curso, tendo como fonte os registros deixados nas diferentes ferramentas disponibilizadas no ambiente virtual. A Avaliação de aprendizagem inclui também avaliação por quizzes.

A versão online do curso a distância contém os questionários e outros instrumentos de avaliação de uma forma simplificada como parte de apresentações, incluindo questões de múltipla escolha, exercícios de arrastar e soltar etc., conforme apresentado no Quadro 8.

Quadro 8 - Possibilidades de atividades avaliativas de forma online

Tipo de atividade avaliativa	Descrição	Exemplos em cursos de ML
<p>Quiz</p>	<ul style="list-style-type: none"> • Permite criar questões flexíveis de múltipla escolha; • Permite criar questão com uma resposta correta; 	 <p>(Machine Learning para todos,2020)</p>  <p>(Machine Learning para todos ,2020)</p>
<p>Arrastar e soltar</p>	<p>Permite criar tarefas de arrastar e soltar com imagens ou textos;</p>	 <p>(ReadyAI, 2019)</p>
<p>Arrastar as palavras</p>	<p>Permite criar tarefas de arrastar e soltar com base em texto;</p>	<p>Aula 5 - Processo de Machine Learning</p>  <p>(Machine Learning para todos,2020)</p>

<p>Redação</p>	<p>Permite criar questão discursiva (pergunta aberta) com feedbacks;</p>	 <p>(Elements of AI, 2019)</p>
-----------------------	--	--

Fonte: Elaborado pela autora

A avaliação é uma etapa importante no processo de ensino aprendizagem, com o objetivo de avaliar o desempenho de alguém ou de algo, busca-se verificar se o que está sendo avaliado demonstra as competências desejáveis. Para aprender, um aluno precisa saber: o que é considerado um bom desempenho em uma determinada atividade; como seu próprio desempenho se relaciona com o bom desempenho; e o que fazer para suprir essa lacuna (STEGEMAN; BARENDSSEN; SMETSERS, 2016). Avaliação baseada no desempenho é realizada com base nos artefatos criados pelos alunos como aprendizagem de resultados. Para cada uma das avaliações baseadas em desempenho, uma rubrica é definida visando um *feedback* para o aluno.

Nas avaliações de desempenho de artefatos criados pelos alunos como resultado da aprendizagem, as rubricas consistem em ferramentas para definição de métricas de avaliação mais transparentes e eficazes (LOBATO et al., 2007). As rubricas permitem um detalhamento mais claro do processo de avaliação e ao mesmo tempo facilitam o diagnóstico de problemas específicos dentro do processo de ensino-aprendizagem (LOBATO et al., 2007, p. 5). Aqui vale ressaltar que a rubrica não é pré-existente, mas sim construída de acordo com os objetivos de aprendizagem.

Uma rubrica é uma ferramenta que auxilia na avaliação do trabalho do aluno. É uma grade de avaliação que consiste em:

- (1) um conjunto de critérios para identificar a característica que deve ser medida;
- (2) uma determinada escala de classificação com vários níveis de desempenho, junto com um sistema de pontuação (qualitativo ou quantitativo) relacionado a cada nível de desempenho;

(3) descritores verbais que detalham textualmente para cada critério os níveis de desempenho.

Para realizar uma correção uniforme e consistente de atividades abertas de *Machine Learning* como parte do ensino de computação costuma-se utilizar rubricas. A avaliação do desempenho do aluno foca na capacidade de aplicação de conceitos aprendidos em contextos de atividades abertas em que não há somente uma resposta correta para determinada atividade (WIGGINS, 1993). Neste tipo de avaliação, o foco é no artefato/resultado final e não em como os resultados foram atingidos. Busca-se verificar se os objetivos de aprendizagem, que definem o que aluno deve ser capaz de fazer ao final de uma Unidade Instrucional, foram atingidos (DRISCOLL; WOOD, 2007).

***Feedback* orientado ao objetivo**

O *feedback* orientado aos objetivos pode ter dois tipos de metas, focada na aprendizagem ou no desempenho (DWECK, 1986; THURLINGS, VERMEULEN, BASTIAENS, & STIJNEN, 2013). A meta focada na aprendizagem é caracterizada pelo desejo do indivíduo em desenvolver uma determinada competência para dominar novas situações/problemas. Já a meta focada no desempenho busca mensurar o desempenho do indivíduo na realização de determinadas atividades em relação a determinados parâmetros, de modo que esta possa ser avaliada e comparada por seus pares (FARR, HOFMANN, & RINGENBACH, 1993; SHUTE, 2007). Entretanto, pesquisas indicam que o *feedback* orientado ao objetivo com foco na aprendizagem tem maior contribuição para aprendizagem do que o foco no desempenho (HOSKA, 1993; NICOL & MCFARLANE-DICK, 2006). O *feedback* focado na aprendizagem auxilia o aluno a compreender que a competência se desenvolve por meio da prática e que o esforço é necessário durante o processo de aprendizagem, e que cometer erros faz parte deste processo (BANGERT-DROWNS, KULIK, KULIK, & MORGAN, 1991; NICOL & MCFARLANEDICK, 2006).

3 ESTADO DA ARTE

Para definir o estado da arte sobre quais modelos de avaliação da aprendizagem do ensino de *Machine Learning* existem na Educação Básica, foi realizado um estudo de mapeamento sistemático seguindo o procedimento proposto por PETERSEN et al. (2008)(2015).

3.1 DEFINIÇÃO DO PROTOCOLO DE REVISÃO

A questão de pesquisa a ser analisada é: quais modelos de avaliação da aprendizagem do ensino de *Machine Learning* existem na Educação Básica?

Para responder essa pergunta, são definidas as seguintes perguntas de análise:

PA1. Quais abordagens de avaliação existem?

PA2. Quais conceitos de *Machine Learning* são avaliados e em quais níveis de aprendizagem?

PA3. Que tipo de avaliação é realizada?

PA4. Que tipo de *feedback* instrucional é apresentado?

PA5. Como as abordagens foram desenvolvidas e avaliadas em termos da confiabilidade e validade?

CrITÉRIOS de inclusão / exclusão: Para esta revisão, foram considerados artigos e/ou materiais em língua inglesa que apresentam Unidades Instrucionais voltadas ao ensino/aprendizagem de ML na Educação Básica e que também apresentam alguma forma de avaliação da aprendizagem do estudante. Artefatos que se concentram no ensino superior/infantil e/ou sem abordar os conceitos de ML foram excluídos. Também foram excluídos estudos que exclusivamente focaram na avaliação das próprias Unidades Instrucionais e não na aprendizagem do aluno. Também foram excluídas publicações como blogs, vídeos ou ferramentas que não fornecem um modelo de avaliação. Ressalta-se que não foram encontrados artigos e/ou materiais em língua portuguesa no contexto da avaliação de aprendizagem de *Machine Learning* na Educação Básica.

CrITÉRIO de qualidade: Foram considerados apenas artigos ou materiais que apresentam informações substanciais sobre o modelo de avaliação de aprendizagem de *Machine Learning*.

Base de dados: São consideradas como fonte de pesquisa os artigos indexados pela ferramenta de busca em bases na área educacional disponíveis na web, por meio das bibliotecas e bancos de dados digitais, incluindo a Biblioteca Digital ACM, IEEEEXplore, SCOPUS (<https://www.scopus.com>) com acesso livre por meio do Portal de periódicos CAPES¹.

Para aumentar a cobertura, também foi usado o GOOGLE, que indexa um grande conjunto de dados em várias fontes diferentes (HADDAWAY et al., 2015), pois nessa área emergente várias unidades de ensino não foram publicadas como artigos científicos. Foi observado também o foco da pesquisa no laboratório de mídia do MIT nessa área, assim também foram buscadas publicações desse grupo de pesquisa. Também foi incluída literatura secundária, descoberta com base na literatura primária encontrada para obter informações mais detalhadas.

Definição da *string* de pesquisa: Como buscas preliminares indicaram a ausência de pesquisas exclusivamente voltadas a modelos de avaliação da aprendizagem de *Machine Learning*, foi direcionada a busca a Unidades Instrucionais voltadas ao ensino de ML para analisá-las em relação às avaliações definidas. Assim, a sequência de busca foi composta por conceitos relacionados à questão de pesquisa, incluindo também sinônimos, conforme indicado no Quadro 9.

¹ Um portal para o acesso de trabalhos científicos, gerenciado pelo Ministério da Educação do Brasil, disponível para instituições autorizadas, como universidades e agências de pesquisa (www.periodicos.capes.gov.br).

Quadro 9 - Palavras-chave

Conceitos principais	Sinônimos
<i>Machine Learning</i>	<i>artificial intelligence, deep learning, data science</i>
<i>K-12</i>	<i>school, kids, teens, children</i>
<i>instructional unit</i>	<i>teach, education, course, MOOC</i>

Fonte: elaborada pela autora

No entanto, pesquisas de testes informais visando a calibração da *string* de pesquisa, não apresentaram resultado relevante em relação aos termos avaliação e classificação, visto que estes artigos retornaram relacionados ao uso de técnicas de *Machine Learning* para análise de aprendizagem ou para apoiar a avaliação de diferentes conhecimentos das áreas. Portanto, esses termos não foram incluídos na sequência de pesquisa com o objetivo de identificar as Unidades Instrucionais e suas avaliações. Como resultado, a *string* de pesquisa foi calibrada e adaptada de acordo com a sintaxe específica da fonte de dados:

(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*).

Quadro 10 - *Strings* de busca para cada fonte

Fonte	Search string
ACM https://dlnext.acm.org/search/advanced	[[Abstract: teach*] OR [Abstract: education] OR [Abstract: course] OR [Abstract: mooc]] AND [[Abstract: "machine learning"] OR [Abstract: "data science"] OR [Abstract: "artificial intelligence"] OR [Abstract: "deep learning"]] AND [[Abstract: "k-12"] OR [Abstract: school*] OR [Abstract: kids] OR [Abstract: children] OR [Abstract: teen*]] AND [Publication Date: (01/01/2010 TO *)]
IEEE https://ieeexplore.ieee.org/search	((("Abstract":teach*) OR ("Abstract":education) OR ("Abstract":course) OR ("Abstract":MOOC) OR AND ("Abstract":“machine learning”) OR ("Abstract":“data science”) OR ("Abstract":“artificial intelligence”) OR ("Abstract":“deep learning”)) AND ((“Abstract”:“k-12”) OR (“Abstract”:school*) OR (“Abstract”:kids) OR (“Abstract”:children) OR (“Abstract”:teen*))) Filters Applied: 2010 - 2020
Scopus https://www2.scopus.com/search	TITLE-ABS-KEY (teach* OR education OR course OR mooc) AND ("machine learning" OR "data science" OR "artificial intelligence" OR "deep learning") AND ("k-12" OR school* OR kids OR children OR teen*)) AND (LIMIT-TO (PUBYEAR , 2020) OR ... OR LIMIT-TO (PUBYEAR , 2010)) AND (LIMIT-TO (SUBJAREA , "COMP"))
Google https://www.google.com/	"machine learning" teach "K-12" school
MIT AI education and media lab https://appinventor.mit.edu/explore/research https://www.media.mit.edu/groups/long-felong-kindergarten/publications/ https://aieducation.mit.edu/#about	--

Fonte: elaborada pela autora

3.2 EXECUÇÃO DA BUSCA

A pesquisa foi realizada em junho de 2020 pela autora (Quadro 11). Várias buscas retornaram muitos resultados, mesmo após uma calibração da sequência de pesquisa. Isso se deve ao fato dos artigos que

descrevem como usar as técnicas de IA na educação, como análise de aprendizado personalizado, correspondem aos mesmos termos de pesquisa. Portanto, mantendo a cadeia de pesquisa, foi limitada a análise apenas às mais relevantes.

Quadro 11 - Número de artigos identificados por repositório e por fase de seleção

Fonte	No. de resultados da busca	No. de resultados analisados	No. de resultados potencialmente relevantes	No. de resultados relevantes
ACM	3.551	200	15	0
IEEE	888	200	8	1
SCOPUS	2.928	200	4	0
Google	37.500.000	500	81	8
MIT media lab	141	14	6	2
Total (sem duplicados)				11

Fonte: elaborada pela autora

Na primeira etapa da análise foram revisados títulos e resumos para identificar trabalhos que correspondessem aos critérios de inclusão, resultando em 114 artefatos potencialmente relevantes. Na segunda etapa, os materiais foram lidos na íntegra para verificar sua relevância com relação aos nossos critérios de inclusão / exclusão. Muitos artigos foram excluídos devido ao seu foco no uso da IA para a educação ou ao “aprendizado profundo” como atividade cognitiva no processo de aprendizado. Também foram excluídos artefatos relacionados a outros estágios educacionais (ensino pré-escolar ou superior) (WILLIAMS et al., 2019a) (WILLIAMS et al., 2019b) (PARK et al., 2019) (BENNETT, 2017) (ESTEVEZ et al., 2019) e os que cobrem a IA, mas não o aprendizado de máquina, por exemplo (CSUNPLUGGED, 2015) (AI4ALL, 2019) (ALI et al., 2019) (PARSONS e SKLAR, 2004) (MIT, 2019). Além disso, foram excluídos materiais que consistem apenas em vídeos explicando *Machine Learning* (CS4FN, 2019) ou ferramentas (AGASSI et al., 2019) (MAKEBLOCK, 2019) ou demos como Google Teachable Machine (GOOGLE, 2017). Artigos que não indicam uso de avaliação também foram excluídos, (FIEBRINK, 2019), (DILLENBOURG e HUAN, 2020). Os duplicados foram eliminados e os artigos que descrevem a mesma Unidade Instrucional foram unificados. Como resultado, 11 Unidades Instrucionais foram consideradas relevantes após a busca, por apresentarem avaliações.

3.3 ANÁLISE DE DADOS

Esta seção apresenta a análise dos dados extraídos das Unidades Instrucionais, de acordo com as questões de pesquisa definidas na Seção 3.1. São apresentadas as abordagens de avaliação encontradas, bem como informações indicando se a abordagem atribui uma nota, e se apresenta *feedback* instrucional.

PA1. Quais abordagens de avaliação existem?

Foram encontradas 11 Unidades Instrucionais que também incluem a avaliação de aprendizagem. As abordagens são voltadas para os Conceitos Básicos de *Machine Learning*, Impactos de ML, Redes Neurais, conforme apresentados no quadro 12.

Quadro 12 - Resumo das avaliações de conceitos e práticas de ML

Nome	Referência	Conceitos de ML	Nível de Aprendizagem	Ciclo Usar-Modificar-Criar	Método de avaliação	tipo de <i>feedback</i>
AI Family Challenge	(AI Family Challenge, 2019)	Conceitos básicos de ML	Lembrar, compreender, aplicar	Usar	Teste: 1 questão de múltipla escolha	Correção automatizada de respostas
AIinSchools	(AIinSchools, 2019)	Conceitos básicos de ML	Lembrar, compreender, aplicar	Criar	Teste: 1 questão de múltipla escolha; Rubrica: 9 critérios em níveis de desempenho de 3 pontos	Correção automatizada de respostas. Avaliação manual por desafio
Apps for good: ML course	(Apps for Good, 2019b)	Conceitos básicos de ML, Impacto do ML	Lembrar, compreender, aplicar	Usar, Criar	Questões de avaliação	Avaliação manual pelo instrutor
AI for Oceans	(Code.org, 2019)	Conceitos básicos de ML	Lembrar, compreender	--	Conclusão da tarefa	Indicação automatizada do grau de conclusão da tarefa, certificado. Nenhuma análise de correção. Feedback de professores para monitorar uma classe
Curiosity Machine - build a neural network	(Curiositymachine, 2019)	Redes neurais	Lembrar, compreender	--	Quiz: 3 multiple-choice questions	Apresentação das respostas dadas, sem correção automática
Elements of AI	(Elements of AI, 2019)	Conceitos básicos de ML, Impacto do ML, Redes neurais	Lembrar, compreender	--	Exercícios com 1-3 questões de texto aberto ou de múltipla escolha	Correção automatizada de respostas de múltipla escolha. Apresentando respostas de exemplo e revisão por pares para respostas de texto. Não

						há nota, mas a quantidade de exercícios realizados é rastreada. Certificado
Alternate Curriculum Unit: AI	(Exploring Computer Science, 2019)	Conceitos básicos de ML, Impacto do ML	Lembrar, compreender	--	Rubricas com critérios de 8 a 14 para a avaliação das apresentações dos alunos	Correção automatizada de respostas
Introduction to ML: Image Classification	(MIT App Inventor, 2019a)	Redes neurais	Lembrar, compreender, aplicar	Usar	Teste de múltipla escolha (2 questões)	Avaliação manual pelo instrutor
Personal Image Classifier	(MIT App Inventor, 2019b)	Redes neurais	Lembrar, compreender, aplicar	Usar	Teste de múltipla escolha (3 questões)	Avaliação manual pelo instrutor
Ready AI AI+Me	(ReadyAI, 2019)	Conceitos básicos de ML, Impacto do ML	Lembrar, compreender	--	Testes de pergunta única (múltipla escolha, arrastar e soltar etc.)	Correção automatizada de respostas Rastreamento de conclusão de tarefa
Kids making AI	(Sakulkeakulsuk <i>et al.</i> , 2018)	Conceitos básicos de ML	Lembrar, compreender	Usar	Gamificação atribuindo pontos de acordo com o desempenho dos modelos de ML criados.	Avaliação manual pelo instrutor

Fonte: elaborada pela autora

As publicações a respeito das unidades instrucionais envolvendo as abordagens de avaliação são recentes. Observa-se que a partir de 2018 foram encontradas as avaliações de conceitos e práticas de *Machine Learning* dentro do contexto educacional, refletindo também a tendência mais recente de ensino de ML/IA, conforme o Quadro 13.

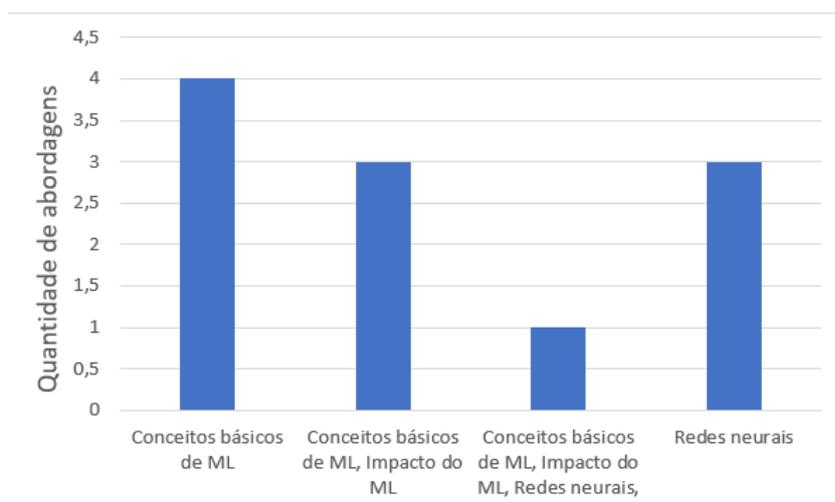
Quadro 13: Ano das publicações das Unidades Instrucionais

Ano de publicação	Unidades Instrucionais
2018	1
2019	10

Fonte: elaborado pela autora

PA2. Quais conceitos de *Machine Learning* são avaliados em que níveis de aprendizagem?

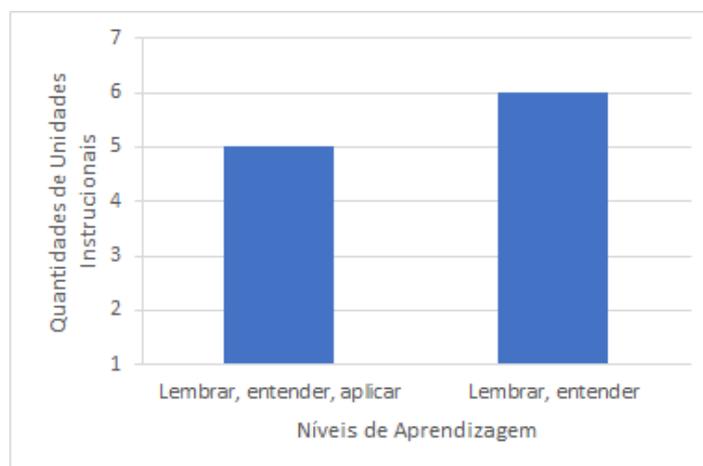
Figura 9 - Visão geral das abordagens com avaliações e práticas de ML



Fonte: elaborado pela autora

Conforme a figura 9, quatro Unidades Instrucionais apresentam conceitos básicos de *Machine learning*, três Unidades Instrucionais apresentam conceitos básicos de *Machine Learning* e Impactos do ML, três Unidades Instrucionais apresentam somente a abordagem em redes neurais, e apenas uma Unidade Instrucional apresenta as três abordagens, a saber: Conceitos básicos de *Machine Learning*, Impactos do ML e Redes Neurais.

Figura 10- Visão geral dos Níveis de aprendizagem



Fonte: elaborado pela autora

Conforme a Figura 10, cinco Unidades Instrucionais aplicam três níveis de aprendizagem em ordem crescente de complexidade com as capacidades de: lembrar (nível factual), entender (ser capaz de lidar com conceitos), aplicar (nível procedimental), e seis Unidades Instrucionais são voltadas a dois níveis de aprendizagem: lembrar e entender, seguindo a taxonomia de Bloom (BLOOM, 1956).

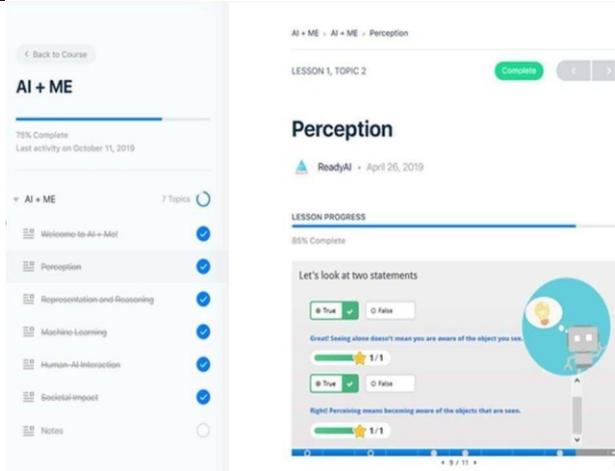
Para facilitar o processo de ensino e aprendizagem, o ciclo *Use-Modify-Create* (LYTLE et al., 2019; LEE et al., 2011) é usado para suportar a progressão da aprendizagem, e podemos observar que cinco Unidades Instrucionais apresentam a primeira etapa (*Use*), e duas Unidades Instrucionais voltadas para a etapa (*Create*).

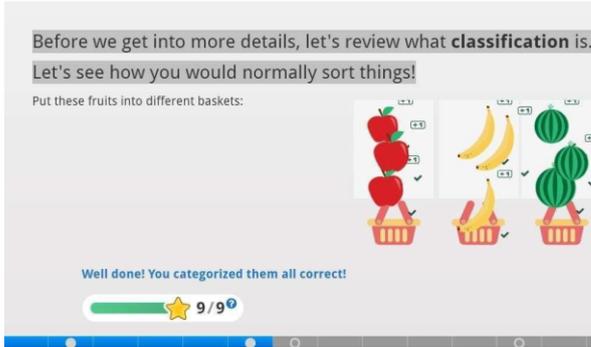
PA3. Tipo de avaliação

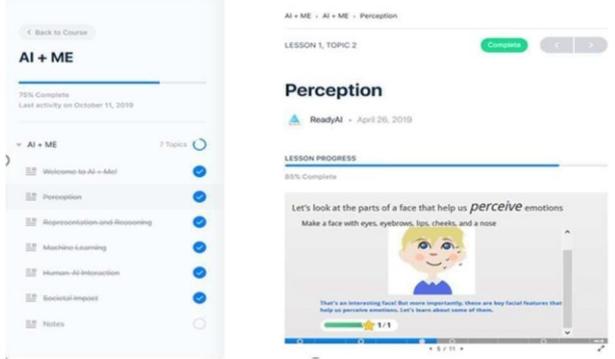
A avaliação de competências no *Machine Learning* com base na análise do método de avaliação é feita usando diferentes formas. A maioria dos cursos *online* usam a avaliação em formato de *quizzes*, sendo elaboradas questões de múltipla escolha, verdadeiro/falso ou respostas curtas. O *quiz* faz parte da avaliação formativa, proporciona maior interação, motivação e auxilia positivamente no processo de ensino e avaliação da aprendizagem dos estudantes (BEZERRA e SANTOS, 2016).

Quadro 14: Exemplos dos tipos de avaliações

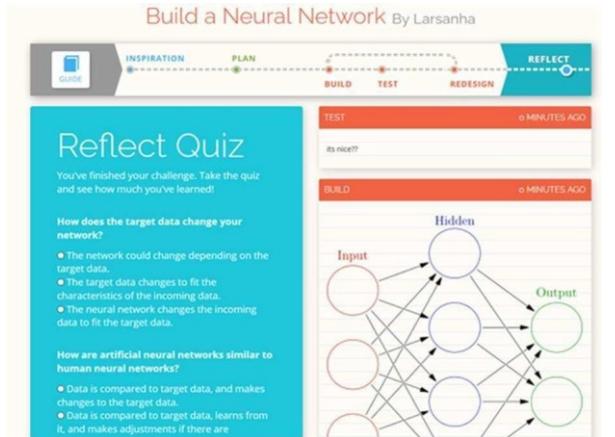
QUIZ

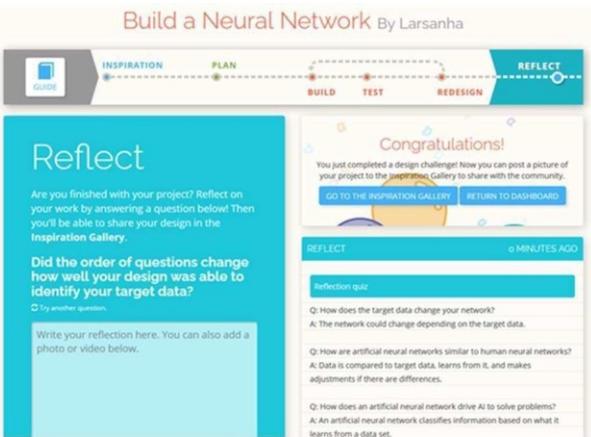






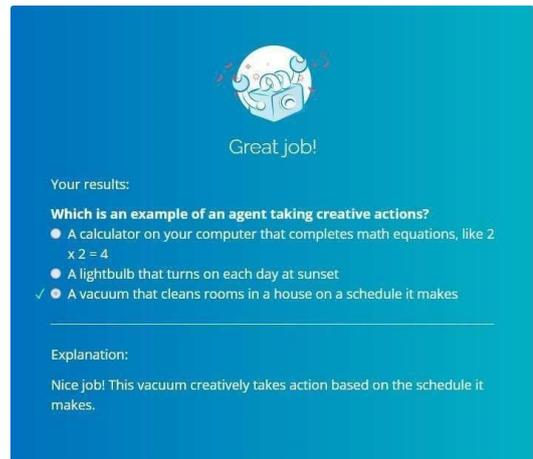
Questão - arrastar e soltar
(READY AI, 2019)



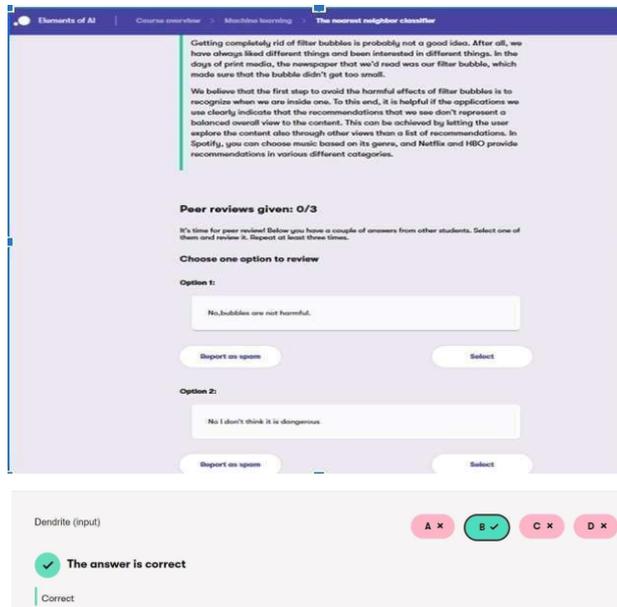




Conclusão de tarefa (CODE.ORG, 2019)



Questões de múltipla escolha (CURIOSITYMACHINE, 2019)



Exercícios com 1-3 questões de texto aberto e de múltipla escolha (ELEMENTS OF AI, 2019)

Rubrica

Final Project Sample Rubric

DO YOU HAVE?	POINTS POSSIBLE	YES	NO	POINTS EARNED
WEBSITE CONTENT				
A home page with an image and a brief description of your topic	5			
3 or more additional pages on your site	15			
Images that support your topic	10			
Cite the source(s) of your images	5			
Complete information for your topic	10			
Cite the source(s) of your information	5			
WEBSITE DESIGN				
Have a background color or image	5			
Incorporate one of the layout styles into your website	5			
Links to all the pages of your website on each page	10			
Integrate navigation bars or menus into your website	5			
Use one shared external .css file for your site	10			
Integrate accessibility features	5			
Storyboard	10			
TOTAL	100			

Rubrica para apresentação do Projeto final dos alunos (Exploring Computer Science, 2019)

Lesson Outline

Theme	Lesson 1 Introduction to Artificial Intelligence	Lesson 2 Introduction to Neural Networks	Lesson 3 The MATHS behind a neural network	Lesson 4 Image Recognition	Lesson 5 Different types of AI	Lesson 6 Assessment
Overview	In this lesson pupils are introduced to the "Smart City" and gain a basic understanding of "Artificial Intelligence". Pupils identify and explore the current use of AI in specific areas such as: <ul style="list-style-type: none"> • Internet Services • Education • Medicine • Media & Entertainment • Helping the Elderly/Disabled • Media & 	In this lesson pupils will be introduced to key terminology around the field of Artificial Intelligence, Machine and Deep Learning. They will gain an overview of the Turing Test and Convolutional Neural Networks.	In this lesson pupils, focus on some of the Mathematics used in a neural network.	In this lesson, pupils will access remote GPUs via a cloud server (Amazon Web Services) to access DIGITS (a front-end tool of common AI frameworks) and use images from the web against those from a known dataset to accurately identify images.	This lesson will get pupils to investigate how AI will change our futures. Areas to be investigated are: <ul style="list-style-type: none"> •Agriculture •Shopping (automated services) •Cooking •Transport including NVIDIA driverless cars (roads) •Medical diagnosis •Assembly line manufacturing 	In this lesson pupils, will Review their learning through completion of a short assessment. Pupils will also collaborate as a class and build a "Wall of fame on women in AI"

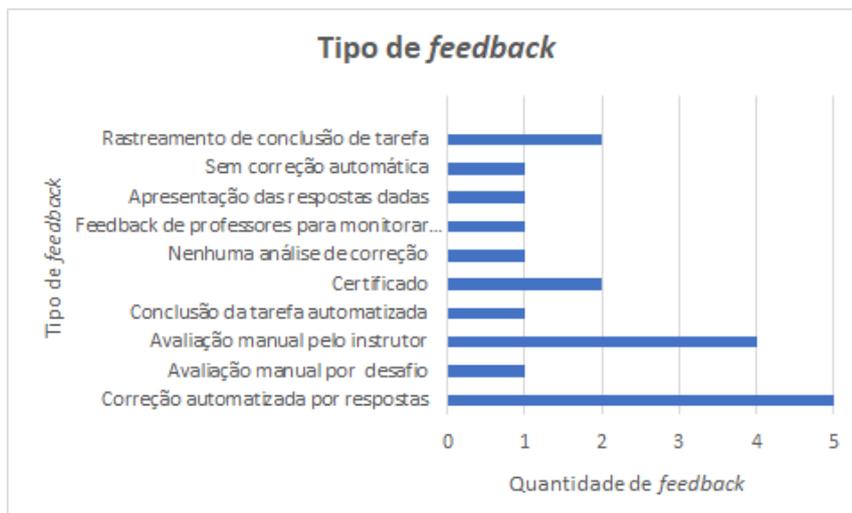
Rubrica para avaliação (AIINSCHOOLS, 2019)

Fonte: Elaborado pela autora

PA4. Que tipo de *feedback* instrucional é apresentado?

O principal objetivo do *feedback* instrucional é auxiliar no desenvolvimento de competências sobre um determinado assunto, identificando nas respostas apresentadas pelo aluno, se estas já se encontram corretas e completas, e caso contrário, indica ao aluno os aspectos que ainda precisam ser corrigidos ou melhorados (RICHARDS & SCHIFFEL, 2005).

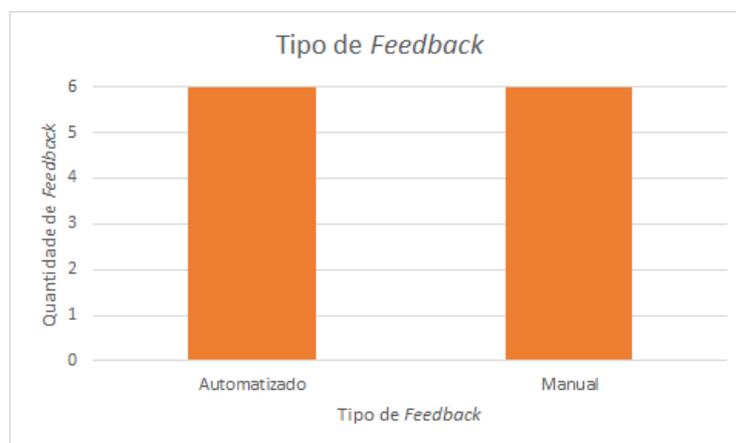
Figura 11 - Tipos de *feedbacks* apresentados nas Unidades Instrucionais



Fonte: elaborado pela autora

O feedback instrucional é um instrumento que pode ser explorado a fim de tornar o processo de aprendizagem mais efetivo. Também quando o feedback instrucional é automatizado como parte de uma ferramenta educacional, o esforço despendido pelo instrutor é reduzido, e sua entrega pode ocorrer de forma imediata aos alunos. Na maioria das avaliações o feedback não é automatizado. Observa-se o feedback somente nos cursos online, e sempre limitada ao tipo de quizzes. Nota-se na Figura 11 que seis tipos de feedback são entregues imediatamente de forma automatizada aos alunos, conforme as interações dos mesmos com a ferramenta, sendo entregue diretamente pela ferramenta de software adotada. Também nota-se que seis tipos de feedback são manuais, sendo estes entregues diretamente pelo instrutor. Não foi encontrado nenhum suporte para avaliação automatizada com base no desempenho/artefato criado pelo aluno.

Figura 12 - Quantidade de *feedback* automatizado e manual



Fonte: elaborado pela autora

Não se preocupando com as atribuições de notas, algumas abordagens utilizam uma pontuação dicotômica, indicando se está certo ou errado, falso ou verdadeiro (READYAI, 2019). A maioria das Unidades Instrucionais não definem a forma de avaliação da aprendizagem dos alunos. Somente AI FAMILY CHALLENGE (2019), AIINSCHOOLS (2019), Exploring Computer Science (2019) e MIT (2019b) propõem uma folha de rubrica/avaliação para analisar as criações dos estudantes. Sakulkueakulsuk et al. (2018) propõe um sistema de pontos com base na acurácia do modelo de ML desenvolvido.

Quadro 15 - Exemplos de avaliações com *feedback* automatizados e manuais

Feedback automatizados / manuais	
<p>Manual</p> <p>6. Assessment</p> <p>Multiple-choice questions</p> <p>Multiple-choice questions assess key concepts of the activity as shown below:</p> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;"> <p>1. If Player1 plays a 30-second long Happy Birthday song, what would happen if you press Button1 and then press Button2 1 second later, using the following blocks?</p>  <p>A. Player1 starts playing a Happy Birthday song. When Button2 is pressed, the song continues to play.</p> <p>B. Player1 starts playing a Happy Birthday song. When Button2 is pressed, the song stops playing.</p> <p>C. Player1 starts playing a Happy Birthday song. When Button2 is pressed, the song continues to play but Button2 plays another copy of the song from the beginning so it sounds like two people are singing with a 1-second delay.</p> <p>D. Player1 starts playing a Happy Birthday song. When Button2 is pressed, the song stops playing and starts over from the beginning.</p> <p>(Answer: D)</p> </div>	<p>Manual</p> <p>7. Assessment</p> <p>Multiple-choice questions to test understanding after the unit</p> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;"> <p>1. You train a model on facial expressions like happy, sad, and surprised with on classifier.appinventor.mit.edu, and build an expression matching game with your model. You play the game and score very highly. You give the app to a friend, and they score very low. Why might that happen?</p> <p>A. The model might need more training to be more accurate.</p> <p>B. The model is biased because you trained it with pictures of yourself.</p> <p>C. Your friend might not be very skilled at making expressions.</p> <p>D. All of the above.</p> <p>(Answer: D. A and B especially ... and C is always a possibility!)</p> </div>

2. Using the following code blocks, and assuming that AnimalLabel.Text starts off blank, what will the text be on AnimalLabel after Button1 is pressed, and then Button2 is pressed?

```

to ChangeLabel newName
do
  set AnimalsLabel .Text to join AnimalsLabel .Text
  get newName
when Button1 .Click
do
  call ChangeLabel
  newName "cow"
when Button2 .Click
do
  call ChangeLabel
  newName "pig"
  
```

- A. cow
 - B. pig
 - C. cow,pig
 - D. ,cow,pig
- (Answer: D)

Avaliação manual pelo instrutor (MIT App Inventor, 2019a)

Manual

3. Your code for an expression matching game includes the following. However, your score never changes from zero. Why are you never getting any score?

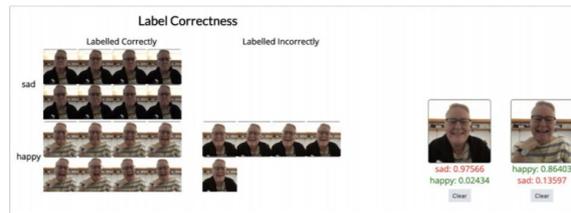
```

when PersonImageClassifier1 .GotClassification
result
do
  if
    select list item list get result = get global currentLabel
    index 1
  then
    set global score to
    select list item list get result * 100
    index 2
  
```

- A. In the set global score block, the index for select list item should be 1, not 2.
- B. You are missing the for each item block, so it is comparing the currentLabel to a list, and it will never be equal.
- C. You should just be comparing if result=currentLabel, and not using select list item.
- D. Your model needs better training.

(Answer: B; the if statement should be inside a for each item in list, and result in the if statement should be changed to item.)

2. You train a model for 2 labels (happy, sad) using classifier.appinventor.mit.edu and get several incorrect labels during the testing phase. Looking at the results below, it labels several "happy" faces incorrectly. What might be the problem?



- A. Not enough test cases.
- B. You need a third label in addition to happy and sad.
- C. All sad images include a black jacket; all happy images do not, so it classifies a black jacket as sad.
- D. There is no problem. You can expect some incorrect labelling during testing.

Answer: C

Manual

3. Assessment questions

- Knowledge – What is machine learning?
- Comprehension – What is the difference between AI and machine learning?
- Application – Give some examples of machine learning being used in day-to-day life.
- Analysis – Why is the use of machine learning increasing rapidly?
- Synthesis – What potential future uses of machine learning can you think of?
- Evaluation – What might be the dangers of using machine learning?

Avaliação manual pelo instrutor (MIT App Inventor, 2019b)

Fonte: elaborada pela autora

3.4 DISCUSSÃO

Como resultado do mapeamento sistemático foram encontradas poucas Unidades Instrucionais voltadas ao ensino de ML que apresentam algum tipo de avaliação. Por ser um tema emergente, a maioria das Unidades Instrucionais foram propostas como unidades extracurriculares, como por exemplo Desafio Família IA (2019). Algumas Unidades Instrucionais levam o aluno a refletir sobre o uso da Inteligência Artificial e os impactos na sociedade atual, e outras abordam as questões morais, variando os conceitos básicos e redes neurais, com aplicação em atividades práticas usando ferramentas e ambientes que facilitem a construção de modelos de *Machine Learning*.

As publicações a respeito das Unidades Instrucionais envolvendo as abordagens de avaliação são recentes, e de 2018 em diante, foram encontradas as avaliações de conceitos e práticas de *Machine Learning* dentro do contexto educacional, refletindo também a tendência mais recente de ensino de ML/IA. Desta forma, a maioria das Unidades Instrucionais são propostas como unidades extracurriculares tanto de forma presencial quanto de cursos *online*. No entanto, como a maioria das Unidades Instrucionais estão disponíveis apenas em inglês, a sua aplicação em outros países é dificultada.

Tendo em vista a importância das abordagens de avaliação das atividades voltadas ao ensino de *Machine Learning* na Educação Básica, não há modelos comumente adotados de como avaliar a aprendizagem de competências de *Machine Learning* na Educação Básica. Porém, observa-se também que a maioria dos modelos não apresentam como foram sistematicamente definidos e/ou avaliados, o que deixa a confiabilidade e validade dessas avaliações questionáveis. No entanto, essa falta de modelos de avaliação, acarreta o processo de aprendizagem do aluno, talvez não fornecendo *feedback* instrucional de forma formativa e somativa para guiar o seu desenvolvimento, sendo importante tanto para o aluno como para o instrutor a fim de orientar o processo de aprendizagem.

Assim, observa-se que a maioria adota uma forma não sistemática de avaliação, principalmente em relação aos níveis de aprendizagem de lembrança e compreensão, com poucas voltadas a avaliar a aprendizagem no nível de aplicação.

As Unidades Instrucionais ensinam competências que vão desde a apresentação do que é ML, até técnicas específicas de ML, bem como os impactos da ML. A Maioria das Unidades Instrucionais focam no estágio de *use*, referente ao ciclo *Use-Modify-Create* (LYTLE et al., 2019; LEE et al., 2011), e poucas avaliações voltadas para a etapa *Modify* ou *Create*.

Foi notável a dificuldade em encontrar registros do planejamento e desenvolvimento das Unidades Instrucionais. Os métodos de avaliação na maioria das Unidades Instrucionais não apresentam como foram sistematicamente desenvolvidas e/ou avaliadas. Desse modo, a aplicação destas Unidades Instrucionais se torna mais complexa, por possuir poucas informações, além da incerteza em relação aos impactos de sua aplicação.

A falta de Unidades instrucionais disponíveis de forma mais ampla em diversas línguas como Português, indica a necessidade de pesquisa e ação nessa área para contribuir com a democratização desse conhecimento também nas Escolas de Educação Básica Brasileiras. Como trabalhos futuros, sugere-se utilizar os resultados desta pesquisa como base para aperfeiçoar o curso *online - Machine Learning* para todos!, desenvolvido pela iniciativa Computação na Escola/INCoS/INE/UFSC, bem como para aperfeiçoar as demais Unidades Instrucionais, além de direções futuras para novos modelos de

avaliações a serem desenvolvidos.

Ameaças à validade da revisão do estado da arte. Como em qualquer mapeamento sistemático, existem algumas ameaças à validade dos resultados. Portanto, foram identificadas ameaças potenciais, e aplicadas estratégias de mitigação para minimizar seus impactos. Os mapeamentos sistemáticos eventualmente irão sofrer o viés de que os resultados positivos são mais prováveis de serem publicados do que os negativos. No entanto, consideramos que as abordagens das Unidades Instrucionais são mais relevantes do que a análise dos resultados, logo consideramos esse risco mínimo. Outro risco é a omissão de algum resultado relevante. Para mitigar esse risco, foi realizada a construção de *string* de busca para ser a mais inclusiva possível, considerando não apenas conceitos básicos, mas também sinônimos. Além disso, considerando a atualidade dos estudos temáticos, também buscamos por Unidades Instrucionais disponíveis *online*, não apenas considerando artigos científicos, a fim de reduzir o risco de exclusão de Unidades Instrucionais existentes. Ameaças à seleção de Unidades Instrucionais relevantes e extração de dados foram moderadas por meio de uma definição detalhada dos critérios de inclusão/exclusão. Foi definido e documentado um protocolo rígido para a seleção dos estudos e todos os autores realizaram a seleção juntos, discutindo até que fosse alcançado um consenso. A extração de dados foi dificultada em alguns casos, pois as informações relevantes para esse estudo nem sempre estavam relatadas explicitamente e, portanto, em alguns casos, tiveram que ser inferidas. No entanto, essa inferência foi feita pela autora e revisada pelo orientador e coorientadora.

4 DESENVOLVIMENTO DO MODELO DE AVALIAÇÃO

Com foco em uma estratégia de aprendizagem ativa levando o aluno a criar modelos de ML, a fim de fornecer uma compreensão prática dos conceitos de ML e redes neurais, e para tornar o ensino de ML envolvente, é desenvolvida uma rubrica que analise o modelo de ML (criado como resultado de aprendizagem) para efeitos de avaliação da aprendizagem. A avaliação é realizada do ponto de vista do instrutor no contexto do estágio de uso no ensino ML. Como atualmente quase todos os alunos são novatos em ML, foi concentrado aqui no estágio de uso do ciclo de aprendizagem. O modelo de avaliação está sendo definido para o seguinte contexto:

Alunos: O público-alvo são alunos da escola pública dos Anos Finais do Ensino Fundamental e do Ensino Médio, com idade entre 12 e 18 anos. De acordo com os conteúdos estabelecidos na BNCC, considera-se que os alunos possuem as competências de analisar, compreender e explicar características, fenômenos e processos relativos ao mundo natural, social e tecnológico (incluindo o digital), criar soluções (inclusive tecnológicas), utilizar diferentes linguagens e tecnologias digitais de informação e comunicação, acessar e disseminar informações, produzir conhecimentos e resolver problemas das Ciências da Natureza de forma crítica, significativa, reflexiva e ética.

De forma geral, os alunos já são alfabetizados na língua nativa (Português do Brasil). Nessa faixa etária estabelecida, possuem uma familiaridade com as línguas estrangeiras (Inglês ou Espanhol), de acordo com os conteúdos estabelecidos na BNCC, porém não necessariamente de forma fluente.

Conforme a Figura 13, o uso de Sistemas Operacionais de celulares de alunos do ensino fundamental corresponde com esse quadro observado de forma geral.

Figura 13 - Distribuição dos sistemas operacionais de celulares no Brasil em 2020



Fonte: Statcounter, fevereiro 2020 - fevereiro 2021

A maioria dos alunos nessa idade possuem conhecimentos prévios e habilidades no uso de computadores e outros dispositivos que acessam a internet.

Figura 14 - Frequência de acesso à internet - Alunos do 9º Ano do Ensino Fundamental

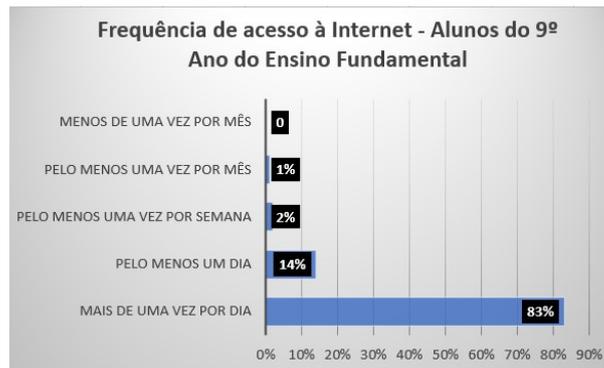
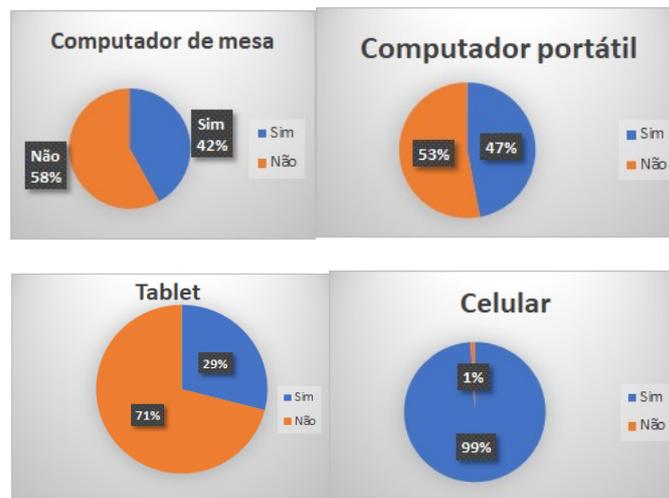


Figura 15 - Dispositivos utilizados para acessar a internet



Fonte: CETIC.BR, 2020, TIC EDUCAÇÃO, 2019

Conforme os dados do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (TIC EDUCAÇÃO, 2019), 83% dos alunos do 9º Ano do Ensino Fundamental das escolas brasileiras acessam a Internet mais de uma vez por dia. E os estudantes dessa faixa etária já fazem uso da internet, celular e outros aparelhos tecnológicos, possuindo habilidades com manuseio e capacidade de aprender a utilizar as ferramentas com facilidade. Sendo que, 99% utilizam o celular para acessar a Internet.

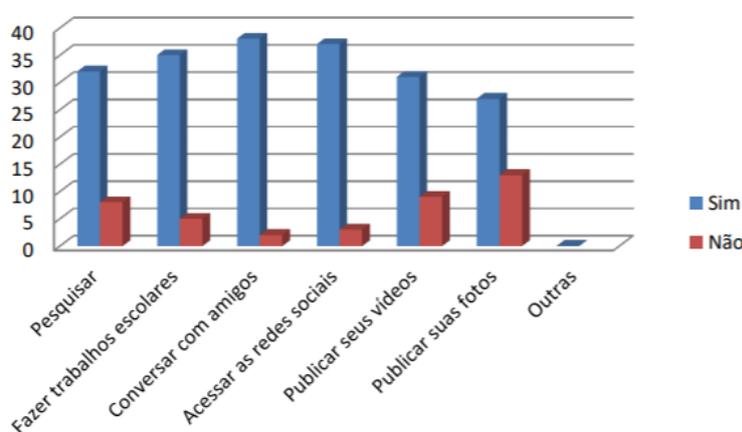
Em relação às habilidades relacionadas à computação, considerando que este tópico começou a ser abordado na última reforma da BNCC, com foco no pensamento computacional, ainda são poucos os alunos que aprenderam esta competência no ensino fundamental, especificamente a programação. Alguns alunos já têm uma compreensão de programas de computador, mas poucos sabem como criar um (FERREIRA et al., 2020).

Com relação a Inteligência Artificial e *Machine Learning*, o currículo proposto pela Sociedade Brasileira de Computação (SBC) sugere que os estudantes tenham contato com esse tema apenas no Ensino Médio (SBC, 2018), com o intuito de que compreendam os fundamentos da inteligência artificial e robótica, sendo que o *Machine Learning* não é citado. Na prática, a abordagem do tema na Educação Básica é

praticamente inexistente no Brasil. Encontram-se relatos do ensino de IA na Educação Básica e no Ensino Médio. Somente no ano de 2020 foi lançado o primeiro curso técnico em Inteligência Artificial do Brasil, o curso visa apresentar ao aluno a capacidade das máquinas de pensarem como seres humanos: Aprender, perceber e decidir quais caminhos seguir, de forma racional diante de determinadas situações (FECAP, 2020).

Quando questionados sobre o que costumam fazer no celular além das atividades realizadas no computador, grande parte dos alunos responderam: Pesquisar, fazer trabalhos escolares, conversar, acessar as redes sociais, publicar seus vídeos e fotos, etc, conforme observado na Figura 16:

Figura 16 - Atividades com as TICs



Fonte: Labrunie, 2017

Ambiente curricular. O ensino da computação no Ensino Fundamental tem sido introduzido nas escolas de forma multidisciplinar ou como atividades extracurriculares (BORDINI, et al., 2016). A computação não faz parte do Currículo Básico no Brasil, apesar da BNCC (MEC, 2018) sugerir competências afins, as Unidades Instrucionais que abordam esse conteúdo geralmente têm curta duração. Atualmente há pouca orientação disponível para o ensino no nível K12, embora haja esforço para atender as necessidades de diretrizes curriculares de Inteligência Artificial. Conforme apresentado no capítulo 2, Fundamentação Teórica, já estão sendo desenvolvidas diretrizes de currículo para ensino de ML, a AI4K12 (TOURETZKY et al., 2019a) e a AI Literacy (LONG & MAGERKO, 2020).

Ambiente escolar: As aulas no modo presencial ocorrem tipicamente em Salas Informatizadas com computadores e acesso à internet. Por exemplo, na Rede Municipal de Florianópolis as Salas Informatizadas são utilizadas por professores de outras áreas não relacionadas a informática (português, matemática, ciências, geografia etc.) com o auxílio do Professor Auxiliar de Tecnologia Educacional para a realização de projetos educativos e trabalhos digitais ensinando alfabetização de TI, como edição

de textos, apresentações, animação (Stop Motion), programação (CODE.ORG), etc.

O total de Escolas na Educação Básica do Brasil é 179.533, sendo que 61.798 escolas possuem laboratório de informática, o que corresponde a 34% do total (Censo Escolar / INEP 2020 - QEdu.org.br).

Em cada Sala Informatizada do Ensino Fundamental da Rede Municipal de Florianópolis, há de 12 a 20 computadores.

Diante dos crescentes casos do Novo Coronavírus (COVID-19) no Brasil e procurando atender às medidas de prevenção e combate ao contágio para preservar a saúde de todos, as atividades nas Unidades Educativas da Rede Municipal de Ensino de Florianópolis foram suspensas na forma presencial, conforme Decretos Municipal e Estadual. Assim as aulas aconteceram de forma remota. Preocupados com a situação, a Prefeitura Municipal de Florianópolis por meio da Secretaria Municipal de Educação, que é responsável pelas duas etapas da Educação Básica: A educação infantil e o ensino fundamental, apresentou o PORTAL EDUCACIONAL, que tem o propósito de estabelecer um diálogo direto com toda a comunidade educativa (Portal Educacional, 2020). Em meados de 2021 as aulas presenciais foram retornadas gradativamente, ocorrendo em grande parte o ensino de forma híbrida, ou seja, parte presencial e parte remota.

O acesso na Rede Municipal foi realizado pelos canais:

- Google Sala de Aula - <https://classroom.google.com>
- Portal Educacional - <https://sites.google.com/prod/sme.pmf.sc.gov.br/portaleducacional/página-inicial>
- Encontros - <https://meet.google.com>

Contexto do curso *Machine Learning* para Todos!

O modelo de avaliação desenvolvido neste trabalho se refere ao curso interativo *on-line Machine Learning* para todos!. O curso *on-line* (MOOC) é voltado para a participação ilimitada e acesso aberto pela web conforme detalhado na seção 2.3. É aplicado de forma assíncrona, sendo de forma individual (sem turma), não necessitando realizar as aulas concomitantemente com demais alunos. Nesse sentido, o estudante pode realizar sua intervenção de resposta participando a qualquer momento, em um período

estipulado previamente ou não, sem necessariamente estar de forma simultânea, presente com outros colegas ou com o professor no ambiente escolhido para interação.

O curso interativo *on-line* também pode ser aplicado com atividades síncronas, no ensino presencial no contexto escolar com instrutor e colegas da turma, ou em plataforma *online*. Nesse contexto, atividades síncronas compreendem-se aquelas que dispõem de um espaço-tempo comum, previamente estabelecido, entre estudante-professor para sua realização, momento esse em que estudantes e professores precisam estar presentes, concomitantemente, na plataforma escolhida para a realização. Com base na análise do domínio do conteúdo e do contexto específico, foi definida a avaliação dos conceitos de ML em conformidade com as diretrizes curriculares da AI4ALL e da literacia em AI definidas por Dung & Magerko (2020), conforme apresentado no capítulo 2.3.

4.2 DEFINIÇÃO DO PLANO DE AVALIAÇÃO

Complementando a avaliação por *quizzes* conforme Quadro 7, o foco do presente trabalho é propor um modelo de avaliação de desempenho do estudante.

Assim foi desenvolvido sistematicamente um modelo de avaliação baseado no desempenho. A rubrica foi revisada e os critérios foram ajustados para o curso *Machine Learning* para todos! a partir de uma análise do contexto educacional da tarefa de aprendizagem (GRESSE VON WANGENHEIM et al., 2021), conforme apresentado no Quadro 16.

Quadro 16 - Rubrica de ML - Estágio de uso

Critérios	Níveis de desempenho		
	Fraco - 0 pt	Aceitável - 1 pt	Bom- 2 pt
Gerenciamento de dados (LO5)			
C1. Quantidade de imagens	Menos de 20 imagens por categoria	21 a 35 imagens por categoria	Mais de 36 imagens por categoria
C2. Relevância de imagens	Muitas imagens não estão relacionadas a tarefa (irrelevantes) e/ou ao menos uma imagem contém conteúdo não ético (violência, nudez, etc)	Ao menos uma imagem é irrelevante, mas não contém imagens não éticas.	Todas as imagens são relacionadas a tarefas de ML e éticas.
C3. Distribuição do conjunto de dados	A quantidade de imagens em cada categoria varia muito. Mais de 10% de variação em ao menos uma categoria (relativo ao total).	A quantidade de imagens entre as categorias tem entre 3% e 10% de variação.	Todas as categorias têm a mesma quantidade de imagens (menos de 3% de variação).

C4. Rotulagem de imagens	Menos de 20% das imagens têm foi rotulado corretamente	Entre 20 e 95% das imagens foram rotuladas corretamente	Mais de 95% das imagens foram rotuladas corretamente
C5. Limpeza de dados	Há várias imagens confusas (fora de foco, vários objetos na mesma imagem, etc.)	Há uma imagem confusa	Nenhuma imagem confusa foi incluída no conjunto de dados
Treinamento de modelo (LO6)			
C6. Treinamento	O modelo não foi treinado	O modelo foi treinado usando parâmetros padrão	O modelo foi treinado com parâmetros ajustados (por exemplo, época, tamanho do lote, taxa de aprendizagem)
Interpretação de desempenho (LO7)			
C7. Testes com novos objetos	Nenhum objeto testado	1-3 objeto testado	Mais de 3 objetos testados
C8. Interpretação de testes	Interpretação errada	(não aplicável)	Interpretação correta
C9. Interpretação da acurácia	Categorias com baixa precisão não são identificadas corretamente e interpretação incorreta em relação ao modelo	Categorias corretamente identificadas com baixa acurácia, mas interpretação incorreta em relação ao modelo	Categorias corretamente identificadas com baixa acurácia e a consequente interpretação a respeito do modelo
C10. Interpretação da matriz de confusão	As classificações errôneas não são identificadas corretamente e a interpretação a respeito do modelo é incorreta	As classificações errôneas foram corretamente identificadas, mas a interpretação a respeito do modelo é incorreta	Identificação correta de erros de classificação e a consequente interpretação com respeito ao modelo
C11. Ajustes /Fez melhorias	Nenhuma nova iteração de desenvolvimento foi relatada	Uma nova iteração com mudanças no conjunto de dados e/ou parâmetros de treinamento foi relatada	Várias iterações com mudanças no conjunto de dados e/ou parâmetros de treinamento foram relatadas

Fonte: Adaptado de Gresse von Wangenheim et al., 2021

5 AVALIAÇÃO DO MODELO

Este capítulo apresenta a avaliação do modelo por meio de uma série de estudos de caso para avaliar os artefatos criados pelos estudantes como resultado da aprendizagem, desenvolvendo um modelo de ML com Google Teachable Machine. Para a avaliação de desempenho desses resultados de aprendizagem é utilizada a rubrica apresentada no Quadro 16.

O objetivo da avaliação da rubrica é voltado à avaliação feita em termos de confiabilidade e validade de construto. A confiabilidade refere-se ao grau de consistência dos itens do modelo, pode ser medida por meio do Coeficiente Ômega (FLORA, 2020). A validade de um construto é a capacidade de medir de fato o que se propõe, incluindo a validade convergente, que é medida pelo grau de correlação entre os itens (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008). Para analisar as intercorrelações entre os itens de uma mesma subdimensão, utilizou-se a correlação policórica, sendo a mais apropriada para variáveis ordinais observadas (OLSSON, 1979).

5.1 DEFINIÇÃO DA AVALIAÇÃO

Objetivo da avaliação será principalmente avaliar a confiabilidade e validade dos instrumentos de avaliação.

O objetivo deste estudo é analisar o modelo de avaliação da aprendizagem do aluno por meio da rubrica para a classificação de imagens utilizando a *Google Teachable Machine*.

A pergunta de pesquisa da avaliação é: O modelo de avaliação da aprendizagem do aluno é confiável e válido?

A qualidade do modelo é avaliada em termos de confiabilidade e validade de construto da rubrica do ponto de vista de pesquisadores no contexto do ensino de computação na Educação Básica. Seguindo a abordagem GQM (BASILI; CALDIERA; ROMBACH, 1994) o objetivo do estudo é definido e decomposto em aspectos de qualidade e questões de análise a serem avaliadas com base nos dados coletados.

Confiabilidade

AQ1: Existe evidência de consistência interna da rubrica para a avaliação da aprendizagem de ML do estudante?

Validade

AQ2: Existe evidência de validade convergente da rubrica para a avaliação da aprendizagem de ML do estudante?

Definição da coleta dos dados.

A partir das perguntas de análise foi definida a operacionalização da coleta de dados. São coletados dados em relação às atividades interativas no curso (*quizzes* etc.) e por meio de relatórios de desenvolvimento de modelo de ML através de formulários *online* apresentando resultados parciais de aprendizagem no processo de desenvolvimento do modelo de *Machine Learning*.

Foi operacionalizada a coleta de dados por meio dos questionários limesurvey (anexo A), obtendo-se assim dados referentes aos resultados de aprendizagem criados pelos estudantes, incluindo o conjunto de dados usado para o treinamento do modelo de ML, parâmetros de treinamento do modelo de ML e resultados da análise e interpretação do desempenho do modelo de ML via formulário *online* no meio do curso.

5.2 APLICAÇÃO E COLETA DE DADOS

O curso ML4ALL foi aplicado em 5 estudos de casos nos anos de 2021 e 2022 a alunos com idade entre 12 e 18 anos, do Ensino Fundamental e Médio, alunos de ambos os sexos, conforme apresentado no Quadro 17.

Antes da aplicação do curso, foram apresentados o objetivo da pesquisa e as questões éticas a todos os participantes e responsáveis. Todos os responsáveis concordaram no termo de consentimento livre e esclarecido (TCLE) e os participantes no termo de assentimento (Anexo - Termos e questionários). A execução do projeto foi aprovada pela CEPSH - Número do parecer: 4.893.560 e pela Secretaria Municipal de Educação de Florianópolis (Anexo - Declaração de parecer do comitê de ética - Secretaria Municipal de Educação de Florianópolis).

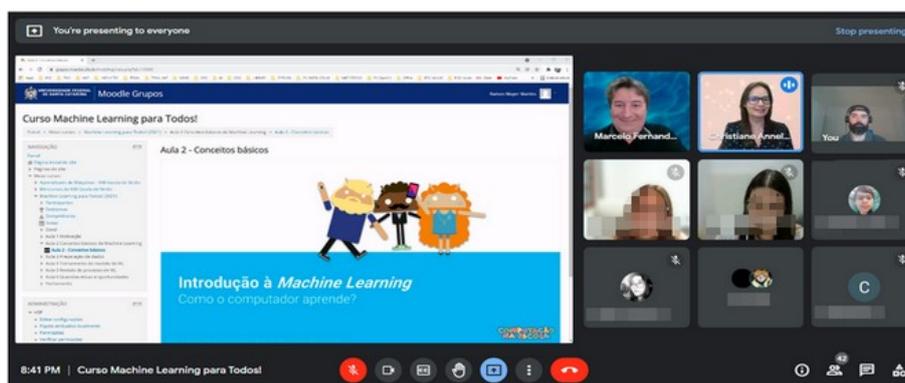
Com exceção de uma aplicação (AP1) em uma escola de Ensino Fundamental em que o curso foi aplicado como parte das aulas, todas as aplicações foram realizadas como atividades extracurriculares.

O curso foi aplicado em diferentes modalidades de ensino, desde aulas presenciais, aulas remotas com instrutores até autoaprendizagem remota.

Quadro 17 - Visão geral das inscrições do curso

Aplicação	Data	Instituição	Modo de instrução	Tipo de atividade	Idade	Estágio Educacional	Nº de estudantes
AP1	Setembro 2021	Escola Básica Municipal Dilma Lúcia dos Santos	Cara a cara	Como parte das aulas da escola	15-16 anos	Ensino Fundamental	12
AP2	Outubro 2021	Instituto Federal Catarinense (IFC)	Acompanhamento remoto do instrutor	Extracurricular	15-17 anos	Ensino Fundamental e Médio	10
AP3	Novembro 2021	Aberto a qualquer estudante interessado organizado pela Universidade Federal de Santa Catarina (UFSC)	Acompanhamento remoto do instrutor	Extracurricular	12-18 anos	Ensino Fundamental e Médio	35
AP4	Março 2022	Aberto a qualquer estudante interessado organizado pela Universidade Federal de Santa Catarina (UFSC)	Acompanhamento remoto do instrutor	Extracurricular	14-18 anos	Ensino Fundamental e Médio	40
AP5	-	Curso online	Remoto como auto-ritmo	Extracurricular	≤ 18 anos	Ensino Fundamental e Médio	11
Total							108

Figura 17 - Alunos durante as aplicações do curso



Fonte: Elaborado pela autora

Os alunos apresentaram voluntariamente os resultados de aprendizagem, bem como o *feedback*. Um total de 108 participantes completaram o curso. A maioria dos participantes eram estudantes do Ensino Médio com mais de 15 anos. Na aplicação AP1 - AP4 mais da metade era do sexo masculino (64%). Mais de 80% dos participantes estavam matriculados na rede pública de ensino brasileira. A seguir, Quadro 18, são demonstrados os dados demográficos dos participantes.

Quadro 18 - Dados demográficos dos participantes

Aplicação	Participantes	Etapas Educacionais		Gênero	
		Ensino Fundamental (≤ 15 anos)	Ensino Médio (>15 anos)	Feminino	Masculino
AP1	12	9 (75.00%)	3 (25.00%)	3 (25.00%)	9 (75.00%)
AP2	10	1 (10.00%)	9 (90.00%)	6 (60.00%)	4 (40.00%)
AP3	35	9 (25.71%)	26 (74.29%)	9 (25.71%)	26 (74.29%)
AP4	40	6 (15.00%)	34 (85.00%)	16 (40.00%)	24 (60.00%)
AP5	11	6 (54.55%)	5 (45.45%)	*	*
Total	108	33 (29.4%)	80 (71.4%)	36 (35.64%)†	65 (64.36%)†

*Informações sobre gênero não foram coletadas como parte do AP5

Durante as aplicações do curso foram coletados os dados dos alunos, por meio de questionários e relatórios sobre o desenvolvimento do modelo de ML. Ao final do curso o *feedback* dos alunos foi coletado por meio de um pós-questionário (Quadro 19).

Quadro 19 - Visão geral dos dados coletados

Aplicação	Nº de relatórios enviados para avaliação baseada em desempenho		
	Relatório 1. C1-C5	Relatório 2. C6	Relatório 3. C7-C11
AP1	12	12	12
AP2	9	9	9
AP3	35	25	35
AP4	31	31	31
AP5	–	–	–
Total	87	77	87

Fonte: Elaborado pela autora

5.3 PREPARAÇÃO DE DADOS

Após a coleta de dados, utilizou-se a linguagem de programação R para analisar os dados coletados. Para preparação dos dados, conforme os objetivos propostos, identificou-se um total de 108 participantes que responderam a rubrica. A maior quantidade de respostas aos itens ocorreu para os itens C7 (Testes com novos objetos) e C8 (Interpretação de testes), em torno de 87 respostas, e a menor participação na faixa de 65 respostas, foi dada ao item C10 (Interpretação de matriz de confusão). Portanto, ficou evidente que muitos estudantes não responderam a todos os itens propostos, comprometendo parte das análises da rubrica.

No início da análise utilizou-se as respostas em sua totalidade, com intuito de realizar uma distribuição de frequências dos 11 itens, conforme demonstrado no Quadro 20:

Quadro 20 - Distribuição das respostas aos itens

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
0 - FRACO	22	1	3	3	17	0	0	34	19	12	13
1 - ACEITÁVEL	16	0	28	44	6	60	3	0	10	8	37
2 - BOM	32	69	39	23	47	10	84	53	37	45	22
TOTAL	70	70	70	70	70	70	87	87	66	65	72

Fonte: Elaborada pela autora

A princípio, por conta da baixa frequência de respostas em algumas categorias, optou-se pelo agrupamento das categorias de alguns itens para as análises complementares. Esclarece-se aqui, que em algumas análises estatísticas, a baixa frequência de respostas nas categorias gera resultados imprecisos, justificando então o agrupamento realizado. Assim, tanto para as análises de correlação, realizada por meio da correlação policórica, que é uma medida específica para itens categorizados, quanto para a consistência das respostas, realizada por meio do Ômega, foram realizados os seguintes filtros e limpezas na base de dados:

a) Eliminadas as respostas dos estudantes que responderam no máximo 3 itens da rubrica;

Observa-se que estudantes que responderam poucos itens foram eliminados da análise, para não gerar resultados imprecisos nas análises estatísticas, considerando que o escore desses alunos seriam muito imprecisos.

b) Os itens C6 e C7 foram recodificados das categorias 2 para 1, e 1 para 0;

Visto não ter respostas na categoria 0 - Fraco, os itens C6 e C7 foram recodificados da categoria 2 - Bom para a categoria 1 - Aceitável, e da categoria 1 - Aceitável para a categoria 0 - Fraco, apenas por conveniência computacional da análise do *software* R.

c) O item C8 foi recodificado da categoria 2 para 1, e mantendo as respostas da categoria 0.

Para melhor análise dos dados, não havendo respostas na categoria 1 - Aceitável, manteve-se as respostas na categoria 0 - Fraco, mas realocadas as respostas da categoria 2 - Bom para a categoria 1 - Aceitável.

d) O item C2 foi eliminado por ter uma alta concentração de resposta em apenas uma categoria.

O item C2 foi eliminado por apresentar apenas uma resposta na categoria 0 - Fraco, e todas as demais respostas na categoria 2 - Bom, ou seja, devido à falta de variabilidade nas respostas desse item, não é possível analisá-lo.

Logo, realizados os filtros e limpezas na base de dados, apresentou-se a seguinte distribuição das respostas, conforme Quadro 21.

Quadro 21 - Distribuição de respostas

	C1	C3	C4	C5	C6	C7	C8	C9	C10	C11
0 - FRACO	22	3	3	17	60	3	31	19	12	9
1 - ACEITÁVEL	16	28	44	6	10	66	38	10	8	32
2 - BOM	32	39	23	47	-	-	-	37	45	19
TOTAL	70	70	70	70	70	69	69	66	65	60

Fonte: Elaborada pela autora

Do total de 108 participantes inicialmente, após a filtragem e limpeza na base de dados, restaram respostas de 79 participantes. A quantidade máxima de respostas por item foi de 70, e a quantidade mínima por item foi de 60 respostas.

5.4 RESULTADOS DA ANÁLISE

5.4.1 Existe evidência de consistência interna da rubrica para a avaliação da aprendizagem de ML do estudante?

Como descrito anteriormente, a consistência interna pode ser analisada por meio do Coeficiente Ômega (Flora, 2020), que é um indicador de consistência interna dos itens de um instrumento. Existem variações do cálculo do Ômega, sendo que há um modelo específico e mais adequado para calcular o Ômega quando os itens apresentam respostas categóricas, como é o caso neste trabalho. De acordo com a literatura, o coeficiente $\hat{\Omega} > 0,70$ indica que os itens da rubrica apresentam boa consistência interna.

Analisando os 10 itens da rubrica para a avaliação da aprendizagem de ML do estudante, obteve-se o valor do Ômega Global 0,646, ficando próximo do mínimo aceitável do Ômega, que é maior que 0,70. No quadro 22 é apresentado o valor do Ômega Global e o valor do Ômega ao eliminar cada item individualmente da rubrica. O cálculo do Ômega com a exclusão do item ajuda a identificar possíveis problemas com o item da rubrica, em suma, se ao excluir o item o valor do Ômega aumentar, isso indica que o referido item não é relevante para ser considerado na rubrica.

Quadro 22 - Coeficiente Ômega - Análise com todos os itens
Excluindo o item

Ômega Global		0,646
C1	Quantidade de imagens	0,547
C3	Distribuição do conjunto de dados	0,488
C4	Rotulagem de imagem	0,582
C5	Limpeza de dados	0,694
C6	Treinamento	0,715
C7	Testes com novos objetos	0,500
C8	Interpretação de testes	0,660
C9	Precisão na interpretação	0,648
C10	Interpretação de matriz de confusão	0,670
C11	Ajustes / Fez melhorias	0,646

Fonte: Elaborada pelo autor

No Quadro 22 observou-se que ao excluir os itens C5, C6, C8, C9 e C10, o coeficiente Ômega aumenta, evidenciando que tais itens não estariam associados ao mesmo traço latente. Ressalta-se que a maior alteração do coeficiente Ômega ocorreu ao excluir o item C6, resultando no valor 0,715. Por outro lado, ao excluir itens C1, C3, C4, C7 o coeficiente Ômega diminui, evidenciando que tais itens estariam associados ao mesmo traço latente.

Esse resultado mostra que pode ter dois blocos de itens na análise, ou seja, a rubrica apresenta duas dimensões, sendo que cada dimensão estaria analisando um traço latente específico. Para explorar um pouco mais esse resultado, na sequência será apresentada a análise do Ômega para cada bloco de itens, conforme resultado encontrado.

Ao realizar a análise da 1ª primeira dimensão com os itens do C1 a C5 (Quadro 23), obteve-se o aumento do Ômega Global para 0,721, sendo o valor acima do mínimo aceitável para uma boa consistência interna. Retirado o item C5- Limpeza de dados, o coeficiente Ômega aumenta para 0,853, apresentando problema nessa dimensão, ou seja, podendo não estar associado ao mesmo traço latente.

Quadro 23 - Coeficiente Ômega - Análise com os itens C1 a C5

Excluindo os itens

Ômega Global		0,721
C1	Quantidade de imagens	0,600
C3	Distribuição do conjunto de dados	0,449
C4	Rotulagem de imagem	0,634
C5	Limpeza de dados	0,853

Fonte: Elaborada pelo autor

Ao analisar a 2ª dimensão, considerando os itens C6 a C11 (Quadro 24), o Ômega Global apresentou o valor de 0,732, também acima do valor esperado para uma boa consistência interna. Apesar de apresentar

um bom coeficiente, foi observado que retirando os itens C6 - Treinamento, C8 - Interpretação de testes, e C10 - Interpretação de matriz de confusão, o valor do Ômega Global aumenta, indicando assim problemas com os demais itens desta dimensão.

Quadro 24 - Coeficiente Ômega - Análise com os itens C6 a C11

Excluindo os itens		Ômega Global
		0,732
C6	Treinamento	0,779
C7	Testes com novos objetos	0,622
C8	Interpretação de testes	0,766
C9	Precisão na interpretação	0,570
C10	Interpretação de matriz de confusão	0,726
C11	Ajustes / Fez melhorias	0,666

Fonte: Elaborada pela autora

5.4.2 Existe evidência de validade convergente da rubrica?

Para obter evidências da validade convergente dos itens da rubrica, calculam-se as intercorrelações dos itens e correlação total do item (DEVELLIS, 2003). Para a validade convergente, espera-se que os critérios da mesma subdimensão tenham correlação mais alta (CARMINES; ZELLER, 1982; TROCHIM; DONNELLY, 2008). Para analisar as intercorrelações entre os itens de uma mesma subdimensão, utilizou-se a correlação policórica, sendo a mais apropriada para variáveis ordinais observadas (OLSSON, 1979), conforme dito anteriormente. O coeficiente de correlação, indica o grau de correlação entre dois itens ordinais (pares de itens). De acordo com Cohen (1998), uma correlação entre os itens é considerada satisfatória, se o coeficiente de correlação for maior ou igual a 0,30, indica que há uma correlação média ou alta entre os itens.

Ao analisar a correlação policórica os itens foram divididos em duas dimensões, sendo que a primeira dimensão foi composta pelos itens de C1 a C5, que estão relacionados ao objetivo de aprendizagem “Gerenciamento de dados”, e a segunda dimensão foi composta pelos itens de C6 a C11, que estão relacionados aos objetivos de aprendizagem “Treinamento do modelo” e “Interpretação de desempenho”, conforme resultados apresentados no Quadro 25.

Quadro 25 - Matriz de correlação policórica dos itens

		C1	C3	C4	C5	C6	C7	C8	C9	C10	C11
Quantidade de imagens	C1	1									
Distribuição do conjunto de dados	C3	0,54	1								
Rotulagem de imagem	C4	0,26	0,67	1							
Limpeza de dados	C5	-0,19	-0,15	-0,13	1						
Treinamento	C6					1					
Testes com novos objetos	C7					-0,24	1				
Interpretação de testes	C8					-0,27	0,51	1			
Precisão na interpretação	C9					0,13	0,47	0,21	1		
Interpretação de matriz de confusão	C10					-0,46	0,01	0,27	0,51	1	
Ajustes / Fez melhorias	C11					0,43	0,17	-0,02	0,55	0,19	1

Fonte: Elaborada pela autora

Observa-se que a primeira dimensão foi apresentada em laranja, e a segunda dimensão foi apresentada em verde. Os pares de itens com maiores correlações foram destacados em negrito. Conforme já mencionado, espera-se que o coeficiente de correlação seja maior ou igual a 0,30. Nessa matriz de correlação policórica foram relacionadas 21 correlações entre os pares, sendo que destas correlações, 7 apresentaram um coeficiente de correlação maior que 0,30, o que corresponde a 33,33% do total. A seguir são demonstradas as maiores correlações dos pares de itens, em cada dimensão, por ordem decrescente:

Quadro 26: Correlação dos pares de itens em cada dimensão, acima do valor esperado de 0,30, por ordem decrescente

Primeira dimensão

Item	Item	Correlação	
C3	C4	0,67	C3 - Distribuição do conjunto de dados e C4 - Rotulagem de imagem
C1	C3	0,54	C1 - Quantidade de imagens e C3 - Distribuição do conjunto de dados

Segunda dimensão

Item	Item	Correlação	
C9	C11	0,55	C9 - Precisão na interpretação e C11 - Ajustes / Fez melhorias
C7	C8	0,51	C7 - Testes com novos objetos e C8 - Interpretação de testes
C9	C10	0,51	C9 - Precisão na interpretação e C10 - Interpretação de matriz de confusão
C7	C9	0,47	C7 - Testes com novos objetos e C9 - Precisão na interpretação
C6	C11	0,43	C6 - Treinamento e C11 - Ajustes / Fez melhorias

Fonte: Elaborada pela autora

Observa-se que o maior coeficiente de correlação na primeira dimensão foi encontrado entre os itens C3 (Distribuição do conjunto de dados) e C4 (Rotulagem de imagem) sendo o valor de 0,67. Na segunda dimensão foi encontrado o maior coeficiente de correlação entre os itens C9 (Precisão na interpretação) e C11 (Ajustes/ Fez melhorias).

Por outro lado, observa-se que o item C5, da primeira dimensão, apenas apresentou correlação negativa ou inversa com os demais itens. A correlação negativa indica que há uma relação inversamente proporcional entre os itens, logo quando um critério aumenta o outro diminui. Quanto à segunda dimensão, observa-se que os seguintes itens apresentaram apenas uma correlação com valor acima de 0,30, a saber: C6 (Treinamento), C8 (Interpretação de dados), e C10 (Interpretação de matriz de confusão). Também ressalta-se a forte correlação entre os itens C6 e C10, no valor de -0,46, que apesar de alta, é negativa, o que justamente não se espera, pois ocorre uma relação inversamente proporcional.

5.4.3 Discussão

Os resultados mostram que a rubrica está perto de alcançar uma confiabilidade aceitável e validade convergente. Percebe-se uma proximidade da confiabilidade aceitável pois o Ômega Global alcançou o valor de 0,646, pouco menor do que o valor mínimo esperado de 0,7.

Com o objetivo de melhorar as análises dos critérios foram estabelecidas duas dimensões, na primeira dimensão estão os critérios C1 a C5, relacionados ao objetivo de aprendizagem “Gerenciamento de dados”, e na segunda dimensão pelos critérios de C6 a C11, que estão relacionados aos objetivos de aprendizagem “Treinamento do modelo”.

Alguns critérios tiveram correlação baixa ou negativa, indicando que podem não estar realmente contribuindo para a avaliação da rubrica. Na primeira dimensão, principalmente o critério C5 (Limpeza de dados), e na segunda dimensão os critérios C6 (Treinamento), C8 (Interpretação de testes), C10 (Interpretação de matriz de confusão), chamam a atenção, pois ao serem excluídos o coeficiente ômega aumenta.

Neste sentido, para que seja melhorada a rubrica, sugere-se uma revisão/ajuste nesses itens para verificar se estão medindo o mesmo traço latente, assim melhorando a correlação nas suas respectivas dimensões. Outra sugestão seria obter mais dados da pesquisa, para então melhorar as análises estatísticas, e realizar outros métodos de análise, como análise fatorial e a Teoria de Resposta ao Item (TRI).

Sugere-se também que seja realizada a avaliação automatizada, com intuito de diminuir os erros na correção da avaliação, e amenizar a sobrecarga dos professores, restando mais tempo para outras atividades/projetos.

Em suma, os resultados da análise demonstram que a rubrica de *Machine Learning* está perto de ser considerada um instrumento confiável e válido para ser aplicado na avaliação de aprendizagem de conceitos de *Machine Learning* na Educação Básica. Ainda assim merece uma atenção no sentido de revisar/ajustar a rubrica, com intuito de melhorar os índices de confiabilidade e validade.

5.4.4 Ameaças à validade

Com intuito de mitigar impactos negativos à validade nesta pesquisa, identificaram-se potenciais ameaças e aplicaram-se métodos para mitigação dos riscos envolvidos. A fim de mitigar as ameaças relacionadas a avaliação da rubrica foi definida e documentada uma metodologia sistemática usando a abordagem GQM (BASILI; CALDIERA; ROMBACH, 1994). A padronização de dados pode ser considerada um risco à qualidade de dados. Como todos os dados foram coletados por meio do curso *Machine Learning* para todos!, o risco foi minimizado. Também para minimizar erros na pontuação da rubrica, a avaliação foi realizada e revisada por diferentes pesquisadores.

A amostra da pesquisa contou com a participação de 108 estudantes, assim pode ser considerada uma ameaça, até mesmo pelo motivo de vários itens ficarem sem respostas. Os estudantes treinaram o modelo, mas alguns estudantes não anexaram o modelo. Notou-se que alguns estudantes não responderam completamente os questionários via limesurvey mesmo sendo reforçado a importância de responder todos os questionários. Assim, a coleta de dados foi prejudicada parcialmente pela coleta incompleta de artefatos criados pelos alunos durante essas atividades práticas e também pode ter comprometido as análises de correlações. Com base nos resultados apresentados na análise, sugere-se um aumento da amostra para analisar e revisar resultados adquiridos no futuro e melhorar as análises estatísticas. Outro risco refere-se à escolha dos métodos estatísticos para análise dos dados. Para mitigar esse risco, realizou-se análise da consistência interna por meio do coeficiente Ômega, e a evidência de validade convergente através da matriz de correlação policórica dos itens.

6 CONCLUSÃO

Os resultados das análises, indicam que a rubrica da avaliação preliminar do desempenho de aprendizagem de *Machine Learning* na Educação Básica apresentou parcialmente a confiabilidade e validade convergente.

Foi avaliada uma rubrica para a avaliação da aprendizagem de conceitos de *Machine Learning*. Assim, os resultados satisfazem os objetivos específicos deste trabalho. Conforme os objetivos específicos, o estado da arte foi analisado por meio de um mapeamento sistemático (Objetivo 1). Os resultados do mapeamento indicaram a falta de abordagens para a avaliação detalhada voltado ao ensino de ML que forneça *feedback* apropriado. Por ser um tema emergente, a maioria das Unidades Instrucionais foram propostas como unidades extracurriculares. A abordagem foi instanciada por meio da rubrica, direcionada à avaliação do modelo de avaliação de competências de ML na Educação Básica (Objetivo 2). A abordagem proposta foi avaliada quanto à confiabilidade e validade convergente por meio de um estudo de caso aplicando a rubrica (Objetivo 3). De forma geral, os itens estão próximos de apresentar a consistência interna esperada, por meio do Coeficiente Ômega. E próximo de obter evidências da validade com valor satisfatório por meio da matriz de correlação policórica. Como trabalhos futuros sugere-se a coleta de uma amostra maior dos dados e a revisão de alguns critérios que chamaram atenção como o C5, C6, C8 e C10, com intuito de fornecer mais subsídios para a análise e contribuir nos diversos critérios da avaliação da rubrica.

Assim, a abordagem do modelo de avaliação contribui para que o ensino de conceitos de *Machine Learning* seja inserido em escolas brasileiras.

REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, N. da C., GRESSE VON WANGENHEIM, C. , HAUCK, J. C. R., BORGATTO, A. F., ANDRADE, D. F. CodeMaster: Um Modelo de Avaliação do Pensamento Computacional na Educação Básica através da Análise de Código de Linguagem de Programação Visual. Anais da X Reunião da ABAVE. São Paulo/SP, 2019.

AI FAMILY CHALLENGE. Página inicial. Disponível em: <<https://www.curiositymachine.org/about/>>, 2019.

AIINSCHOOLS. Página inicial. Disponível em: <<http://aiinschools.com/>>, 2019.

AI4K12. AI4K12 Big Idea #3 Progression Chart - Working Draft of Big Ideia 3, 2020.

APPS FOR GOOD. Página inicial. Disponível em: <<https://www.appsforgood.org/courses/machine-learning/>>, 2019.

AMAZON. Amazon Machine Learning, AWS Documentation. <https://docs.aws.amazon.com/machine-learning/latest/dg/building-machinelearning.html>, 2019.

BASE NACIONAL COMUM CURRICULAR (BNCC). Disponível em: <http://basenacionalcomum.mec.gov.br/>, 2018.

BASIL, V. R.; CALDIERA, G.; ROMBACH, H. D. Goal Question Metric Paradigm. 2 ed. MARCINIAK, J. J. Encyclopedia of Software Engineering, John Wiley & Sons, 1994.

BLACK, P., & WILIAM, D. Assessment and classroom learning. Assessment in Education: Principles, Policy, and Practice, 5(1), 7–74, 1998.

BLACK, P., & WILIAM, D. Inside the black box: Raising standards through classroom assessment. London: King’s College, School of Education, 1998.

BORDINI, A. et al. Computação na Educação Básica no Brasil: o Estado da Arte. RITA, 23(2), 2016.

BRANCH, R. Instructional design: The ADDIE approach. 2nd ed. Athens: Springer, 2009.

CARMINES, E. G.; ZELLER, R. A. Reliability and validity assessment. 5ª ed. Beverly Hills: Sage Publications Inc, 1982.

CETIC, Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação. CGI.BR/NIC.BR. Disponível em: <https://Cetic.br>, 2020.

CNE.. Iniciativa Computação na Escola, 2019.

CODE.ORG. Página inicial. Disponível em: <<https://curriculum.code.org/hoc/plugged/9//>>, 2019.

CURIOSITYMACHINE. Página inicial. Disponível em: <<https://www.curiositymachine.org/challenges/126/>>, 2019.

CSTA. “Computer Science Framework/Standards”, 2016/2017.

DE BEM, A. B., LANZER, E. A., TAMBOSI E., SANCHEZ, O. P., BERNARDI, P. Validade e confiabilidade de instrumento de avaliação da docência sob a ótica dos modelos de equação estrutural. Avaliação: Revista da Avaliação da Educação Superior (Campinas), v. 16 (2), p. 375-401, 2011.

- DEVELLIS, R. F. Scale development: theory and applications. Thousand Oaks: SAGE Publications, 2003.
- DEVELLIS, R. F. Scale Development: Theory and Applications. Beverly Hills: SAGE Publications, 2016.
- DRISCOLL, A., WOOD, S. Developing Outcomes-Based Assessment for Learner-Centered Education: a Faculty Introduction. Sterling, VA, USA: Stylus Publishing, 2007.
- DWECK, C. Motivational processes affecting learning. *American Psychologist*, 41(1), pp. 1040–1048, 1986.
- ELEMENTS OF AI, 2019. Página inicial. Disponível em: <<https://course.elementsofai.com/4>>. Acesso em: 10 de junho de 2020.
- EXPLORING COMPUTER SCIENCE, 2019. Página inicial. Disponível em: <<http://www.exploringcs.org/for-teachers-districts/artificial-intelligence>>. Acesso em: 10 de junho de 2020.
- FARR, J., HOFMANN, D., & RINGENBACH, K. Goal orientation and action control theory: Implications for industrial and organizational psychology. In: L. COOLER, *International review of industrial and organizational psychology* (pp. 193–232). New York: Wiley, 1993.
- FECAP- Fundação Escola de Comércio Álvares Penteado. Inteligência Artificial chega à grade do ensino técnico do Brasil. Notícia veiculada na home page da FECAP, em 13 de janeiro de 2020.
- FERREIRA, M. N. F. Ensinando Design de Interface de Usuário de Aplicativos Móveis no Ensino Fundamental, 2020.
- FILATRO, Andréa. *Design Instrucional Contextualizado: Educação e Tecnologia*. São Paulo: SENAC, 2004.
- FILATRO, Andréa. *Design Instrucional na Prática*. S.P.: Pearson Education, 2008.
- FLORA, D. B., Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates. Department of Psychology, York University, 2020.
- GIJSELAERS, W. H. Connecting problem based practices with educational theory. In: L. W. & W. H. G. (Eds.), ed. *Bringing problem-based learning to higher education: Theory and practice*. San Francisco: Jossey-Bass, p. 13-21, 1996.
- GIL, A. C. *Como elaborar projetos de pesquisa*. 5 ed. São Paulo: Atlas, 2010.
- GOODFELLOW, I. *Deep Learning*. MIT Press, Cambridge, MA, Disponível em <<http://www.deeplearningbook.org>>, 2016.
- GRESSE VON WANGENHEIM, C., HAUCK, J. C. R.; PACHECO, F. S., BERTONCELI BUENO, M. F. *Visual Tools for Teaching Machine Learning in K-12: A Ten-Year Systematic Mapping*, 2021.

- GRESSE VON WANGENHEIM, C., MARQUES, L. S., & HAUCK, J. C. R. Machine Learning for All – Introducing Machine Learning in K-12, <https://doi.org/10.31235/osf.io/wj5ne>, 2020.
- Haddaway, N. R. et al. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PloS one*, 10(9), 2015.
- HOSKA, D. Motivating learners through CBI feedback: Developing a positive learner perspective. In: V. DEMPSEY, *Interactive instruction and feedback* (pp. 105-132). Englewood Cliffs: Educational Technology Publications, 1993.
- H5P. Create, share and reuse interactive html5 content in your browser. Disponível em: <<https://h5p.org/>>. Acesso em: 29 de outubro de 2020.
- HURWITZ, J., KIRSCH, D. *Machine Learning For Dummies*, IBM Limited Edition, 2018.
- LEE, I. et al. Computational thinking for youth in practice. *ACM Inroads*, 2(1), 32–37, 2011.
- LOBATO, A. S. et al. Uma rubrica para avaliação de cursos de programação centrada em avaliação automática. SBIE - Workshop de Ambientes de apoio à Aprendizagem de Algoritmos e Programação, 2007.
- LONG, D.; MAGERKO, B. What is AI Literacy? Competencies and Design Considerations. Proc. of the CHI Conference on Human Factors in Computing Systems., Honolulu, HI, USA, 2020.
- 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.
- LYTLE, N., DONG, YIHUAN., BOUNAJIM, DOLLY. Use, Modify, Create: Comparing Computational Thinking Lesson Progressions for STEM Classes. In *Innovation and Technology in Computer Science Education (ITiCSE '19)*, July 15–17, Aberdeen, Scotland Uk. ACM, New York, NY, USA, Article , 7 pages. <https://doi.org/10.1145/3304221.3319786>, 2019.
- MARQUES, L. S., GRESSE VON WANGENHEIM, C. Desenvolvimento de um Tutorial online para Ensinar Aprendizado de Máquina no Ensino Fundamental II. Relatório Técnico, Iniciativa Computação na Escola/INCoD/INE/UFSC, Florianópolis, 2020.
- MARQUES, L. S., GRESSE VON WANGENHEIM, C. Introdução à Machine Learning Desenvolvimento de uma unidade instrucional para ensino de ML na educação básica. Relatório Técnico, Iniciativa Computação na Escola/INCoD/INE/UFSC, Florianópolis, 2020.
- MARQUES, L. S.; GRESSE VON WANGENHEIM, C.; HAUCK, J. C. R. Teaching Machine Learning in School: A Systematic Mapping of the State of the Art. *Informatics in Education*, 2020.
- MARQUES, S. L., GRESSE VON WANGENHEIM, C. Introdução à Machine Learning Desenvolvimento de uma unidade instrucional para ensino de ML na educação básica. Relatório Técnico, Iniciativa Computação na Escola/INCoD/INE/UFSC, Florianópolis, 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.
- MIT APP INVENTOR. Página inicial. Disponível em: <<http://appinventor.mit.edu/explore/resources/ai/image-classification-look-extension>>, 2019.
- MIT APP INVENTOR, 2019b. Página inicial. Disponível em: <<http://appinventor.mit.edu/explore/resources/ai/personal-image-classifier>>. Acesso em: 10 de junho de 2020.

MITCHELL, T. M.. Machine Learning. McGraw-Hill Education, New York, ModelZoo. Disponível em: < <https://modelzoo.co>>, 1997.

MOSKAL, B. M.; LEYDENS, J. A. Scoring Rubric Development: Validity and Reliability. Practical Assessment, Research & Evaluation, v. 7, n. 10, 2000.

NICOL, D., & MCFARLANE-DICK, D. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in higher education, 31(2), pp. 199–218, 2006.

OLSSON, U. Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika, v. 44, n. 4, p. 443–460, 1979.

PELLE. R. Desenvolvimento de um analisador de código para snap! voltado ao ensino de computação na educação básica. <https://repositorio.ufsc.br/handle/123456789/187855>, 2018.

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, 2008/2015.

PORTAL EDUCACIONAL (google.com), 2020.

PORTAL QEDU. Disponível em: <https://academia.qedu.org.br/>. Acesso em 15 de julho 2020.

READYAI, 2019. Página inicial. Disponível em: <<https://edu.readyai.org/courses/aime/>>. Acesso em: 10 de junho de 2020.

RICHARDS, M., & SCHIFFEL, J. A distance learning framework for automatic instructor replies: articulable tacit knowledge used for feedback upon request. In Proc. of the IEEE SoutheastCon, (pp. 611- 620). Lauderdale/USA, 2005.

RIPLEY, B. Pattern Recognition and Neural Networks, University Press, Cambridge, 2008.

ROYAL SOCIETY. Machine learning: the power and promise of computers that learn by example. Technical Report, 2017.

SAKULKUEAKULSUK. Kids making AI: Integrating Machine Learning, Gamification, and Social Context in STEM Education, 2018.

SBC. Disponível em: <<https://www.sbc.org.br/educacao/diretrizes-para-ensino-de-computacao-na-educacao-basica>>, 2019

SHUTE, V. Focus on formative *feedback*. Princeton: Educational Testing Service (ETS), 2007.

SHUTE, V. J., Hansen, E. G., & Almond, R. G. An assessment for learning system called ACED: Designing for learning effectiveness and accessibility (ETS Research Rep. No.RR-07-26). Princeton, NJ: ETS, 2007.

SKYMIND AI. Disponível em: < <https://skymind.ai/wiki/deep-reinforcement-learning>>, 2019.

STATCOUNTER. Mobile Operating System Market Share in Brazil. Disponível em: <<https://gs.statcounter.com/os-market-share/mobile/brazil>>, 2020.

- SURIF, J., IBRAHIM, H. N.; DALIM, S. F. Problem Solving: Algorithms and Conceptual and Open-ended Problems in Chemistry. *Procedia - Social and Behavioral Sciences*, v.116, p. 4955-4963, 2014.
- STEGEMAN, M.; BARENSEN, E.; SMETSERS, S. Designing a rubric for feedback on code quality in programming courses. *Proceedings of the 16th Koli Calling International Conference on Computing Education Research*, p. 160-164, 2016.
- THARWAT, A. Classification assessment methods. *Applied Computing and Informatics*, in press, 2019.
- 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.
- TEACHABLE MACHINE. Train a computer to recognize your own images, sounds, & poses. 2021. Disponível em: < <https://teachablemachine.withgoogle.com>>. Acesso em: Setembro de 2021.
- THURLINGS, M., VERMEULEN, M., BASTIAENS, T., & STIJNEN, S. Understanding feedback: A learning theory perspective. *Educational Research Review*, 9(1), pp. 1-15, 2013.
- TIC EDUCAÇÃO. Pesquisa sobre o uso das tecnologias de informação e comunicação nas escolas brasileiras, 2019.
- TISSENBAUM, M.; SHELDON, J.; ABELSON, H. From Computational Thinking to Computational Action. From computational thinking to computational action. *Communications of the ACM*, 62(3), p.39, 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.
- HITRON, T., YOAV ORLEV, IDDO WALD, ARIEL SHAMIR, HADAS EREL, and OREN ZUCKERMAN. Can Children Understand Machine Learning Concepts? The Effect of Uncovering Black Boxes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Paper 415, 1–11. DOI:<https://doi.org/10.1145/3290605.3300645>, 2019.
- TOURETZKY, D. S., GARDNER MCCUNE, C., MARTIN, F., SEEHORN, D. K12 Guidelines for Artificial Intelligence: What Students Should Know. *Proc. of the ISTE Conference*, Philadelphia, PA, USA. <https://riu.cead.unb.br/orientacoes/2-publicacoes/76-sincrono-assincrono>, 2019.
- TROCHIM, W. M.; DONNELLY, J. P. *Research methods knowledge base*. 3^a ed. Mason, OH: Atomic Dog Publishing, 2008.
- WIGGINS, G. P. *The Jossey-Bass education series. Assessing student performance: Exploring the purpose and limits of testing*. San Francisco: Jossey-Bass, 1993.
- YANG, M.; THUNG, G. Classification of Trash for Recyclability Status. Disponível em <http://cs229.stanford.edu/proj2016/report/ThungYang-ClassificationOfTrashForRecyclabilityStatus-report.pdf>. Acesso em 10 maio, 2020.
- YIN, R. K. *Case study research: design and methods*. Ed. 2, Thousand Oaks: SAGE Publications, 2001.

ANEXO - TERMOS E QUESTIONÁRIOS

Termo de Consentimento Livre e Esclarecido - TCLE

(disponível como formulário online <https://forms.gle/PpgU4SbnSbekLeE29>)

Termo de assentimento

(disponível como formulário online junto com TCLE <https://forms.gle/PpgU4SbnSbekLeE29>)

Declaração de parecer do comitê de ética - Secretaria Municipal de Educação de Florianópolis

SECRETARIA MUNICIPAL DE EDUCAÇÃO
DIRETORIA DE GESTÃO ESCOLAR
GERÊNCIA DE FORMAÇÃO CONTINUADA
Rua Ferreira Lima, 82 – Centro
CEP 88014-420 – Florianópolis – SC
Telefones: (48) 32120922 – (48) 32120923

Florianópolis, 05 de julho de 2021.

DECLARAÇÃO

Declaro para os devidos fins e efeitos legais que, objetivando atender as exigências para a obtenção de parecer do Comitê de Ética em Pesquisa com Seres Humanos, e como representante legal da Secretaria Municipal de Educação de Florianópolis (Gerência de Formação Continuada), tomei conhecimento do projeto de pesquisa: “AVALIAÇÃO DE APRENDIZAGEM DE CONCEITOS DE MACHINE LEARNING”, em desenvolvimento no Métodos e Gestão em Avaliação – PPGMGA, da Universidade Federal de Santa Catarina (UFSC), nível de mestrado, no período de 2021. A pesquisadora Abisague Belem Garcia está sob orientação do Prof. Dr Adriano F. Borgatto Cumprirei os termos das Resoluções CNS nº 466/2012, nº 510/2016 e suas complementares, e como esta instituição tem condição para o desenvolvimento deste projeto, autorizo a sua execução nos termos propostos.



Fabrícia Luiz Souza - Diretora
Diretoria de Gestão Escolar
Matrícula 15829-1

Questionário demográfico

(disponível como formulário online <https://forms.gle/PpgU4SbnSbekLeE29>)

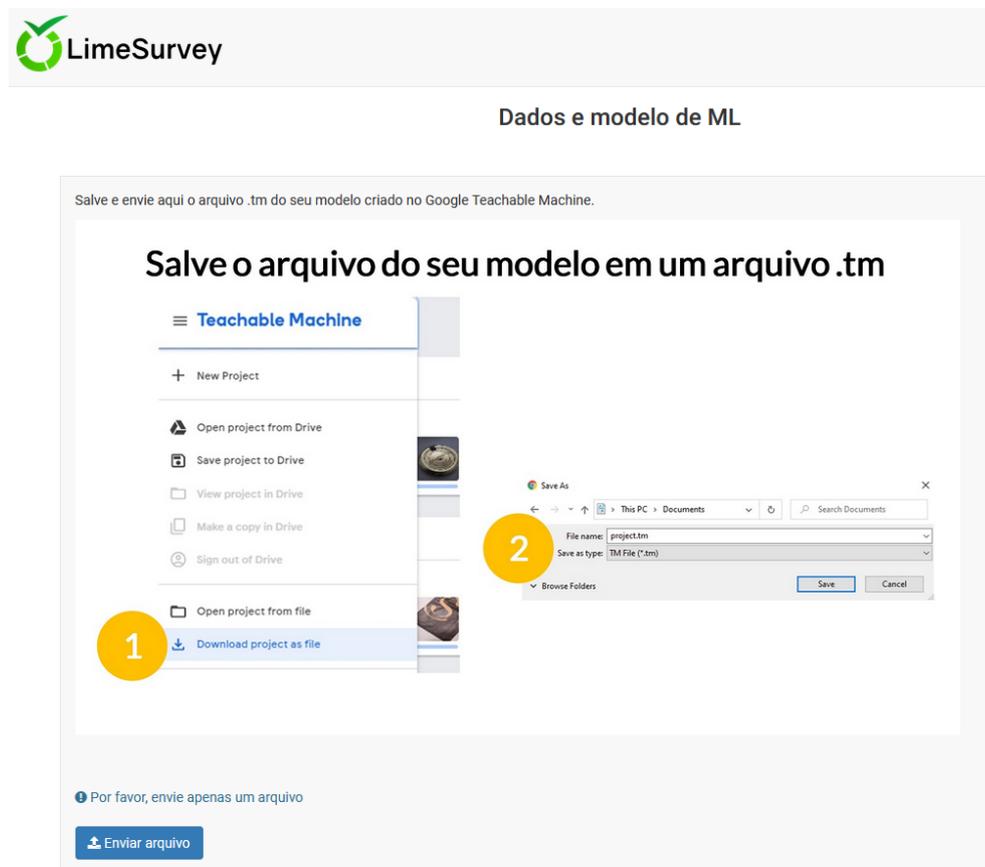
Questionários referentes aos resultados de aprendizagem

Questionarios limesurvey

(disponível como formulários online

<https://survey-gqs.ufsc.br/index.php/697285?lang=pt-BR> <https://survey-gqs.ufsc.br/index.php/994382?lang=pt-BR>
<https://survey-gqs.ufsc.br/index.php/187679?lang=pt-BR> <https://survey-gqs.ufsc.br/index.php/152191?lang=pt-BR> <https://survey-gqs.ufsc.br/index.php/819688?lang=pt-BR>)

coletando os resultados parciais durante o desenvolvimento do modelo de ML



Salve e envie aqui o arquivo .tm do seu modelo criado no Google Teachable Machine.

Salve o arquivo do seu modelo em um arquivo .tm

Teachable Machine

- + New Project
- Open project from Drive
- Save project to Drive
- View project in Drive
- Make a copy in Drive
- Sign out of Drive
- Open project from file
- 1 Download project as file

Save As

File name: project.tm

Save as type: TM File (*.tm)

2 Save Cancel

Por favor, envie apenas um arquivo

Enviar arquivo

(disponível como formulário online <https://survey-gqs.ufsc.br/index.php/697285?lang=pt-BR>)

Relate os testes que você fez com novos objetos.

	Categoria de lixo do objeto				Categoria indicada pelo modelo			
	Metal	Papel	Plástico	Vidro	Metal	Papel	Plástico	Vidro
Objeto novo 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Objeto novo 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Objeto novo 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Objeto novo 4	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Objeto novo 5	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Quantas vezes o modelo acertou?

Escolha uma das seguintes respostas:

Por favor, selecione... ▾

Por favor, selecione...

0

1

2

3

4

5

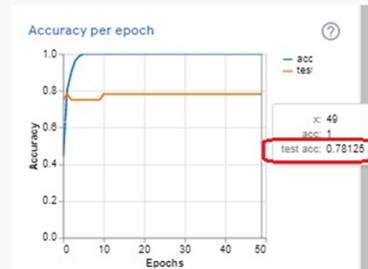
Estes resultados indicam que?

Escolha uma das seguintes respostas:

- o modelo está funcionando perfeitamente
- o modelo precisa ser melhorado

(disponível como formulário online <https://survey-gqs.ufsc.br/index.php/994382?lang=pt-BR>)

Qual o valor geral de acurácia do seu modelo?



Apenas números podem ser usados nesse campo.

A sua resposta deve ser entre 0.00 e 1.00

Analisando a tabela de acurácia, você pode observar, que

Escolha uma das seguintes respostas:

- todas as categorias estão sendo reconhecidas com uma acurácia acima de 90%
- alguma(s) categorias são reconhecidas com acurácia abaixo de 90%, quais?

Isto indica o que?

Escolha uma das seguintes respostas:

- o modelo está funcionando perfeitamente
- o modelo precisa ser melhorado, como?

Por favor, coloque aqui o seu comentário:

Qual é a acurácia por categoria?

CLASS	ACCURACY	# SAMPLES
vidro	0.50	4
papel	0.75	4
metal	0.60	5
plástico	1.00	5

Apenas números podem ser preenchidos nestes campos.

Cada resposta deve ser entre 0.00 e 1.00

Metal

Papel

Plástico

Vidro

Por favor, coloque aqui o seu comentário:

(disponível como formulário online <https://survey-gqs.ufsc.br/index.php/187679?lang=pt-BR>)

Faça o upload da imagem da matriz de confusão do seu modelo.

Salve a imagem da matriz de confusão



Class	metal	paper	plastic	glass
metal	6	1	0	1
paper	0	7	1	0
plastic	0	0	6	2
glass	1	0	0	1

Clique com o botão direito do mouse em cima da imagem

Save image as...
Copy image

Por favor, envie apenas um arquivo

Analisando a matriz de confusão, você pode observar, que:

Comentar apenas quando você selecionar uma resposta.

todas as imagens do conjunto de teste estão sendo classificadas corretamente

alguma(s) das imagens de metal são erradamente classificadas como

alguma(s) das imagens de papel são erradamente classificadas como

alguma(s) das imagens de plástico são erradamente classificadas como

alguma(s) das imagens de vidro são erradamente classificadas como

Isto indica o que?

Escolha uma das seguintes respostas:

o modelo está funcionando perfeitamente

o modelo precisa ser melhorado, como?

Por favor, coloque aqui o seu comentário:

(disponível como formulário online <https://survey-gqs.ufsc.br/index.php/152191?lang=pt-BR>)

Você tentou melhorar o desempenho do seu modelo?

Escolha uma das seguintes respostas:

não

sim, uma vez, mudando o que? Melhorou?

sim, várias vezes, mudando o que? Melhorou?

Por favor, coloque aqui o seu comentário:

(disponível como formulário online <https://survey-gqs.ufsc.br/index.php/819688?lang=pt-BR>)

Questionário referente a experiência de aprendizagem

(disponível como formulário online <https://forms.gle/FwxM5AooVKVu6v9w8>).