



Ministério da Educação
Secretaria de Educação Profissional e Tecnológica
Instituto Federal Catarinense
Campus Videira

THALES FERREIRA BATISTA

**ENSINO PERSONALIZADO DE MATEMÁTICA:
OPORTUNIDADES E TÉCNICAS COMPUTACIONAIS**

Videira
2026

THALES FERREIRA BATISTA

**ENSINO PERSONALIZADO DE MATEMÁTICA:
OPORTUNIDADES E TÉCNICAS COMPUTACIONAIS**

Trabalho de Conclusão de Curso apresentado ao Curso de graduação em Ciência da Computação do Instituto Federal Catarinense – Campus Videira para obtenção do título de bacharel em Ciência da Computação.

Orientador: Prof. Dr. Rafael Zanin (IFC)

Coorientador: Prof. Dr. Manassés Ribeiro (IFC)

Videira

2026

THALES FERREIRA BATISTA

**ENSINO PERSONALIZADO DE MATEMÁTICA:
OPORTUNIDADES E TÉCNICAS COMPUTACIONAIS**

Este Trabalho de Curso foi julgado adequado para a obtenção do título de Bacharel em Ciência da Computação e aprovado em sua forma final pelo curso de graduação em Ciência da Computação do Instituto Federal Catarinense – Campus Videira.

Videira (SC), ____ de _____ de 2026

Prof. Dr. Rafael Zanin
Orientador
Instituto Federal Catarinense – Campus Videira

Prof. Dr. Manassés Ribeiro
Coorientador
Instituto Federal Catarinense – Campus Videira

BANCA EXAMINADORA

Prof. Me. Fabio Pinheiro
Instituto Federal Catarinense – Campus Videira

Prof. Dr. Dani Prestini
Instituto Federal Catarinense – Campus Videira

Dedico este trabalho a...

AGRADECIMENTOS

Agradeço primeiramente...

*“A educação é a arma mais poderosa
que você pode usar para mudar o mundo.”*

(Nelson Mandela)

RESUMO

Este trabalho investigou a aplicação de técnicas computacionais ao ensino de matemática por meio de uma revisão sistemática da literatura e do desenvolvimento de um protótipo de sistema inteligente para diagnóstico de competências. A revisão, cujo relato seguiu as diretrizes PRISMA 2020, identificou 9.431 registros em quatro bases científicas, dos quais 17 estudos foram incluídos após processo rigoroso de triagem e avaliação de qualidade metodológica (MMAT 2018). Os resultados evidenciaram a predominância de técnicas de *machine learning* supervisionado (76,5%), com destaque para algoritmos de classificação como *Random Forest* e SVM, aplicados principalmente à predição de desempenho (52,9%) e personalização de trajetórias de aprendizagem (17,6%). Com base nessas evidências, foi desenvolvido um protótipo computacional fundamentado em *Random Forest* e SVM, com módulo de explicabilidade (SHAP/LIME), alinhado às competências da BNCC e voltado ao apoio de professores no diagnóstico individualizado de competências matemáticas. O trabalho contribuiu com uma síntese reproduzível da literatura, um *pipeline* automatizado de revisão sistemática e um protótipo funcional que integra predição de desempenho com explicabilidade para o contexto educacional brasileiro.

Palavras-chave: ensino de matemática. *machine learning*. revisão sistemática. diagnóstico de competências. BNCC.

ABSTRACT

This work investigated the application of computational techniques to mathematics education through a systematic literature review and the development of an intelligent system prototype for competency diagnosis. The review, reported following PRISMA 2020 guidelines, identified 9,431 records across four scientific databases, of which 17 studies were included after rigorous screening and methodological quality assessment (MMAT 2018). Results showed the predominance of supervised machine learning techniques (76.5%), particularly classification algorithms such as Random Forest and SVM, primarily applied to performance prediction (52.9%) and personalized learning path optimization (17.6%). Based on this evidence, a computational prototype was developed using Random Forest and SVM with an explainability module (SHAP/LIME), aligned with Brazil's BNCC curriculum standards and designed to support teachers in individualized mathematical competency diagnosis. The work contributed a reproducible literature synthesis, an automated systematic review pipeline, and a functional prototype integrating performance prediction with explainability for the Brazilian educational context.

Key-words: mathematics education. machine learning. systematic review. competency diagnosis. BNCC.

LISTA DE ILUSTRAÇÕES

Figura 1 – Fluxo PRISMA 2020 da revisão sistemática.	36
Figura 2 – Funil de seleção de estudos.	37
Figura 3 – Registros identificados por base de dados (inclui duplicatas) (n = 9.431). . .	38
Figura 4 – Distribuição de técnicas de IA nos estudos incluídos.	48
Figura 5 – Distribuição temporal dos estudos incluídos (2016–2025).	49
Figura 6 – Distribuição das pontuações de relevância dos estudos incluídos.	50
Figura 7 – Arquitetura em três camadas do protótipo.	59

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
ACO	<i>Ant Colony Optimization</i> (Otimização por Colônia de Formigas)
ANN	<i>Artificial Neural Networks</i> (Redes Neurais Artificiais)
BNCC	Base Nacional Comum Curricular
CNN	<i>Convolutional Neural Networks</i> (Redes Neurais Convolucionais)
DL	<i>Deep Learning</i> (Aprendizado Profundo)
DT	<i>Decision Trees</i> (Árvores de Decisão)
EDM	<i>Educational Data Mining</i> (Mineração de Dados Educacionais)
IA	Inteligência Artificial
IFC	Instituto Federal Catarinense
IOM	<i>Institute of Medicine</i>
KNN	<i>K-Nearest Neighbors</i> (K-Vizinhos Mais Próximos)
LA	<i>Learning Analytics</i> (Análítica de Aprendizagem)
LIME	<i>Local Interpretable Model-agnostic Explanations</i>
LR	<i>Logistic Regression</i> (Regressão Logística)
ML	<i>Machine Learning</i> (Aprendizado de Máquina)
MMAT	<i>Mixed Methods Appraisal Tool</i>
MVP	<i>Minimum Viable Product</i> (Produto Mínimo Viável)
NB	<i>Naive Bayes</i>
NLP	<i>Natural Language Processing</i> (Processamento de Linguagem Natural)
PICOS	<i>Population, Intervention, Comparison, Outcomes, Study Design</i>
PISA	<i>Programme for International Student Assessment</i>
PRISMA	<i>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</i>
PTC	Projeto de Trabalho de Conclusão de Curso

RF	<i>Random Forest</i> (Floresta Aleatória)
RL	<i>Reinforcement Learning</i> (Aprendizado por Reforço)
SAEB	Sistema de Avaliação da Educação Básica
SHAP	<i>SHapley Additive exPlanations</i>
SMOTE	<i>Synthetic Minority Over-sampling Technique</i>
STEM	<i>Science, Technology, Engineering, Mathematics</i>
STI	Sistemas Tutores Inteligentes
SVM	<i>Support Vector Machines</i> (Máquinas de Vetores de Suporte)
TCC	Trabalho de Conclusão de Curso
XAI	<i>eXplainable Artificial Intelligence</i> (Inteligência Artificial Explicável)

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Justificativa	15
1.2	Problema de Pesquisa	16
1.3	Objetivos	17
1.3.1	Objetivo Geral	17
1.3.2	Objetivos Específicos	17
1.4	Estrutura do Trabalho	17
1.4.1	Fase 1: Revisão Sistemática da Literatura	18
1.4.2	Fase 2: Desenvolvimento do Protótipo	18
1.4.3	Fase 3: Validação Experimental	18
1.4.4	Organização do Documento	19
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	Contexto e Relevância	20
2.2	Ensino de Matemática: Desafios e Perspectivas	20
2.2.1	Desafios do Ensino de Matemática no Brasil	20
2.2.2	Teorias de Aprendizagem Relevantes	21
2.2.2.1	Construtivismo de Piaget	21
2.2.2.2	Zona de Desenvolvimento Proximal de Vygotsky	22
2.2.2.3	Aprendizagem Significativa de Ausubel	22
2.2.2.4	Convergência com a Aprendizagem Personalizada e Adaptativa	22
2.2.3	Tecnologia e Ensino de Matemática	23
2.3	Técnicas Computacionais na Educação Matemática	23
2.3.1	Machine Learning e Inteligência Artificial	23
2.3.2	<i>Learning Analytics e Educational Data Mining</i>	24
2.3.3	Sistemas Tutores Inteligentes	24
2.3.4	Aprendizagem Adaptativa e Personalizada	25
2.4	Avaliação Automatizada e Métricas de Desempenho	26
2.4.1	Tipos de Avaliação	26
2.4.2	Métricas Computacionais	26
2.5	Contribuições da Fundamentação	27

3	METODOLOGIA	28
3.1	Protocolo e Relato da Revisão Sistemática	28
3.2	Estratégia de Busca	28
3.2.1	Bases de Dados e APIs	28
3.2.2	Estratégia de Busca Bilíngue	29
3.3	CrITÉrios de Seleção (PICOS)	30
3.3.1	Justificativa do Recorte Temporal	31
3.3.2	CrITÉrios de Inclusão	31
3.3.3	CrITÉrios de Exclusão	31
3.4	Processo de Seleção (PRISMA)	32
3.4.1	Identificação	32
3.4.2	Triagem	32
3.4.3	Elegibilidade	32
3.4.4	Inclusão	33
3.5	Deduplicação	33
3.6	Infraestrutura Tecnológica	33
3.6.1	Pipeline Automatizado	33
3.6.2	Reprodutibilidade	34
3.7	Avaliação da Qualidade Metodológica	34
4	REVISÃO SISTEMÁTICA DA LITERATURA	36
4.1	Fluxo PRISMA 2020	36
4.2	Estatísticas Descritivas	37
4.3	Síntese dos Estudos Incluídos	38
4.4	Qualidade Metodológica dos Estudos Incluídos	42
4.4.1	Síntese da Qualidade Metodológica	45
4.5	Análise Temática	46
4.5.1	Termos Mais Frequentes	46
4.5.2	Categorias Temáticas Emergentes	46
4.5.3	Distribuição por Abordagem Técnica e Finalidade	47
4.6	Resultados de Eficácia Reportados	49
4.6.1	Magnitude de Efeito	49
4.6.2	Métricas de Avaliação Empregadas	50

4.7	Limitações Identificadas	50
4.7.1	Limitações Técnicas	50
4.7.2	Limitações Pedagógicas	51
4.7.3	Limitações Metodológicas	51
4.7.4	Limitações Éticas	51
4.8	Limitações da Revisão Sistemática	51
4.8.1	Limitações de Processo	52
4.8.2	Limitações da Evidência	52
4.9	Mapeamento de Lacunas e Direcionamento para Fase 2	52
4.9.1	Lacunas Técnicas	53
4.9.2	Lacunas Pedagógicas	53
4.9.3	Direcionamento para Fase 2	53
4.10	Desenvolvimento e Validação	54
4.10.1	Fase 2 — Desenvolvimento de Protótipo (fevereiro–julho/2026)	54
4.10.2	Fase 3 — Validação Experimental (julho–novembro/2026)	54
4.10.3	Síntese Final do TCC (outubro–novembro/2026)	54
5	DESENVOLVIMENTO DO PROTÓTIPO	55
5.1	Fundamentação das Decisões Técnicas	55
5.1.1	<i>Random Forest</i> como Algoritmo Primário	55
5.1.2	SVM como Algoritmo Secundário	56
5.1.3	Métricas de Explicabilidade (XAI)	56
5.2	Requisitos do Protótipo	57
5.2.1	Requisitos Funcionais	57
5.2.2	Requisitos Não Funcionais	58
5.3	Arquitetura do Sistema	58
5.3.1	Camada de Dados	59
5.3.2	Camada de Processamento ML	59
5.3.3	Camada de Apresentação	60
5.4	Bases de Dados e Fontes	60
5.4.1	Microdados SAEB/INEP	60
5.4.2	Microdados PISA	61
5.4.3	Geração de Dados Sintéticos	61

5.4.4	Dados de Escola Parceira	61
5.5	Algoritmos e Implementação	61
5.5.1	<i>Random Forest</i>	61
5.5.2	Support Vector Machine	62
5.5.3	Engenharia de <i>Features</i>	62
5.5.4	Métricas de Avaliação	62
5.5.5	Componente de Explicabilidade (XAI)	63
5.6	Resultados Preliminares do Protótipo	63
6	RESULTADOS E DISCUSSÃO	65
6.1	Resultados da Validação do Protótipo	65
6.1.1	Descrição do Experimento	65
6.1.2	Desempenho dos Algoritmos	65
6.1.3	Análise de Explicabilidade	65
6.2	Discussão	65
6.2.1	Comparação com a Literatura	65
6.2.2	Implicações para a Prática Docente	65
6.2.3	Limitações do Protótipo	65
7	CONCLUSÃO	66
7.1	Síntese das Contribuições	66
7.2	Limitações do Trabalho	66
7.3	Trabalhos Futuros	66
8	CRONOGRAMA	67
	REFERÊNCIAS	68
APÊNDICE A	CHECKLIST PRISMA 2020	71

1 INTRODUÇÃO

O ensino de matemática enfrenta o desafio constante de atender às diversas necessidades de aprendizagem dos alunos em salas de aula heterogêneas. A personalização do ensino, embora reconhecida como uma abordagem eficaz para melhorar o desempenho e o engajamento dos estudantes, é uma tarefa complexa e demorada para os professores (ZHANG, 2023; ZHANG; ZHU; FENG, 2025). A dificuldade em diagnosticar com precisão e em tempo hábil as competências e dificuldades individuais de cada aluno representa uma barreira significativa para a otimização dos planos de ensino (APPIAH-ODAME, 2024).

Nesse contexto, as tecnologias computacionais — como *machine learning*, análise de dados educacionais (*learning analytics*) e inteligência artificial — surgiram como aliadas poderosas (USKOV et al., 2019; TJAHYADI, 2025). Essas técnicas ofereceram o potencial de automatizar a avaliação diagnóstica e fornecer aos educadores informações estruturadas e acionáveis para intervenções pedagógicas mais direcionadas e eficazes (PEJIC; MOLCER; GULAČI, 2021; SOKKHEY et al., 2020). Estudos recentes demonstram que sistemas adaptativos baseados em dados podem melhorar significativamente o desempenho estudantil e reduzir o tempo necessário para alcançar objetivos de aprendizagem (JOSE et al., 2024).

1.1 JUSTIFICATIVA

A relevância deste projeto reside no seu potencial de impacto tanto para educadores quanto para alunos. Para os professores, a síntese sistemática das técnicas computacionais aplicadas ao ensino de matemática oferece fundamentação científica sólida para a tomada de decisão sobre investimentos em tecnologia educacional e adoção de práticas baseadas em evidências (DEPREN; AŞKIN; ÖZ, 2017). Além disso, a identificação de abordagens eficazes pode orientar o desenvolvimento de ferramentas que reduzam a carga de trabalho manual de avaliação e forneçam *insights* acionáveis baseados em dados, permitindo que os educadores dediquem mais tempo à elaboração de estratégias pedagógicas personalizadas (MERTASARI; SASTRI; PASCIMA, 2023).

Para os alunos, um ensino adaptado às suas necessidades individuais — fundamentado em diagnósticos automatizados precisos — pode resultar em melhoria significativa na aprendizagem, na motivação e na redução da ansiedade em relação à matemática (JOSE et al., 2024). Estudos reportam ganhos de aprendizagem de 10 a 20% em sistemas adaptativos quando

comparados a abordagens tradicionais, além de redução de até 20% no tempo necessário para atingir objetivos pedagógicos (ZHANG; ZHU; FENG, 2025).

Do ponto de vista acadêmico, a literatura sobre IA educacional encontra-se fragmentada em diversas áreas, carecendo de consenso terminológico. O relato estruturado conforme as diretrizes PRISMA 2020 buscou garantir rigor, transparência e reprodutibilidade na comunicação dos achados desta revisão.

Praticamente, este trabalho ofereceu aos professores de matemática evidências empíricas para a adoção de tecnologias educacionais, além de fundamentar cientificamente o desenvolvimento do protótipo nas fases subsequentes do TCC, garantindo escolhas técnicas e pedagógicas respaldadas por dados (ÜNAL, 2020; MILIĆEVIĆ; MARINOVIĆ; JEFTIĆ, 2024).

O acompanhamento individualizado do progresso dos alunos em matemática é fundamental para um ensino de qualidade. No entanto, professores frequentemente carecem de ferramentas eficientes para diagnosticar as competências específicas de cada estudante em larga escala (KUMAR et al., 2022), o que dificulta a adaptação dos planos de ensino às necessidades reais da turma. A ausência de um sistema automatizado para essa finalidade leva a um ensino mais generalizado, que pode não atender adequadamente nem os alunos com dificuldades nem os mais avançados.

Apesar do crescente número de pesquisas sobre inteligência artificial aplicada ao ensino de matemática, a literatura científica apresenta-se fragmentada e dispersa em múltiplas bases de dados, disciplinas e abordagens metodológicas. Pesquisadores e desenvolvedores enfrentam dificuldades para identificar quais técnicas computacionais são mais adequadas para diferentes objetivos pedagógicos e quais abordagens foram efetivamente validadas em contextos educacionais reais (MACLELLAN, 2017).

1.2 PROBLEMA DE PESQUISA

Diante desse contexto, o presente trabalho aborda o seguinte problema de pesquisa:

Como identificar e sintetizar, através de revisão sistemática da literatura, as principais técnicas computacionais aplicadas ao ensino de matemática, de modo a fundamentar o desenvolvimento de uma ferramenta que auxilie professores na otimização de seus planos de ensino a partir da identificação automatizada das competências e dificuldades individuais dos alunos?

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Mapear e analisar sistematicamente as aplicações de técnicas computacionais — especialmente *machine learning*, *learning analytics* e sistemas de tutoria inteligente — no contexto do ensino de matemática, identificando tendências, lacunas de pesquisa e oportunidades para o desenvolvimento de um modelo computacional (MVP) que auxilie professores na personalização do ensino e no diagnóstico de competências.

1.3.2 Objetivos Específicos

- **OE1:** Realizar revisão sistemática da literatura, com relato estruturado conforme PRISMA 2020 (PAGE et al., 2021), para identificar estudos que apliquem técnicas computacionais no ensino de matemática, publicados nos últimos 10 anos (2016-2025).
- **OE2:** Identificar e categorizar as principais abordagens de inteligência artificial (*Machine Learning*, *Deep Learning*, *NLP*, *Educational Data Mining*) aplicadas ao ensino de matemática reportadas na literatura científica.
- **OE3:** Classificar as aplicações identificadas segundo suas finalidades pedagógicas: tutoria inteligente, diagnóstico de dificuldades, avaliação automatizada, personalização de conteúdo, predição de desempenho e *feedback* adaptativo.
- **OE4:** Analisar criticamente as metodologias de avaliação utilizadas para validar a eficácia de sistemas computacionais em contextos educacionais, identificando boas práticas e limitações metodológicas.
- **OE5:** Mapear sistematicamente as lacunas de pesquisa, limitações técnicas e desafios reportados nos estudos incluídos, propondo direções prioritárias para o desenvolvimento de ferramentas computacionais eficazes no ensino de matemática.
- **OE6:** Criar um *pipeline* automatizado e reproduzível para coleta, processamento e análise de literatura científica, contribuindo para futuras revisões sistemáticas no campo e fundamentando o desenvolvimento do protótipo nas fases subsequentes do TCC.

1.4 ESTRUTURA DO TRABALHO

Este Trabalho de Conclusão de Curso (TCC) apresenta os resultados de um projeto de pesquisa dividido em três etapas sequenciais:

1.4.1 Fase 1: Revisão Sistemática da Literatura

A primeira fase realizou uma revisão sistemática da literatura cujo relato seguiu as diretrizes PRISMA 2020 para mapear o estado da arte das técnicas computacionais aplicadas ao ensino de matemática. Esta fase foi fundamentalmente **exploratória e analítica**, tendo como objetivo:

- Identificar quais técnicas de IA, *machine learning* e *learning analytics* têm sido aplicadas no ensino de matemática
- Analisar os resultados reportados e metodologias de avaliação utilizadas
- Mapear lacunas de pesquisa e oportunidades de desenvolvimento
- Criar uma base de conhecimento estruturada para orientar as fases seguintes

Entrega realizada: Relatório de revisão sistemática contendo: protocolo de pesquisa, análise de 17 estudos incluídos (de 9.431 identificados, 6.914 únicos após deduplicação), síntese narrativa das aplicações, mapeamento de lacunas e diretrizes para desenvolvimento de ferramentas educacionais.

1.4.2 Fase 2: Desenvolvimento do Protótipo

A Fase 2 consistiu no **projeto e implementação** de um protótipo funcional de ferramenta computacional para diagnóstico de competências matemáticas, fundamentado nos achados da revisão sistemática. Atividades realizadas:

- Levantamento de requisitos funcionais e não-funcionais baseado na literatura
- Definição de arquitetura de *software* e escolha de tecnologias
- Implementação de algoritmos de *machine learning* para diagnóstico automatizado
- Desenvolvimento de interface para professores e alunos
- Integração com bases de dados educacionais e sistemas de gestão escolar

1.4.3 Fase 3: Validação Experimental

A Fase 3 consistiu na **validação empírica** do protótipo desenvolvido em contexto educacional real. Atividades realizadas:

- Planejamento de estudo experimental (*design*, amostra, instrumentos)

- Coleta de dados em ambiente escolar controlado
- Análise quantitativa e qualitativa dos resultados
- Avaliação de eficácia, usabilidade e aceitação pelos usuários
- Refinamento do protótipo com base nos resultados

1.4.4 Organização do Documento

O presente documento está organizado da seguinte forma:

Capítulo 1 - Introdução: Apresentou a contextualização do tema, o problema de pesquisa, os objetivos, a justificativa e a estrutura do trabalho (este capítulo).

Capítulo 2 - Fundamentação Teórica: Apresentou os conceitos fundamentais sobre Inteligência Artificial, *Machine Learning*, *Learning Analytics*, Educação Matemática e Revisões Sistemáticas que embasaram teoricamente a pesquisa.

Capítulo 3 - Metodologia: Descreveu em detalhes a metodologia de revisão sistemática adotada, com relato conforme PRISMA 2020, incluindo estratégias de busca bilíngue (72 consultas em 4 APIs), critérios de seleção PICOS, processo de triagem e extração de dados.

Capítulo 4 - Revisão Sistemática da Literatura: Apresentou os resultados da revisão sistemática, incluindo fluxo PRISMA (9.431 identificados → 6.914 únicos → 17 incluídos), síntese dos estudos incluídos (Tabela 1), análise temática das aplicações de IA identificadas, e mapeamento de lacunas de pesquisa.

Capítulo 5 - Desenvolvimento do Protótipo: Descreveu as decisões técnicas, a arquitetura e a implementação do protótipo de sistema inteligente para diagnóstico de competências matemáticas, fundamentado nas evidências da revisão sistemática.

Capítulo 6 - Cronograma: Apresentou o planejamento temporal detalhado das três fases do projeto (Fase 1: março-novembro 2025; Fase 2: fevereiro-julho 2026; Fase 3: julho-novembro 2026).

Referências Bibliográficas: Lista completa das obras citadas ao longo do trabalho, formatadas segundo normas ABNT NBR 6023:2018, incluindo os 17 estudos incluídos na revisão sistemática.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 CONTEXTO E RELEVÂNCIA

A transformação digital tem impactado significativamente diversas áreas do conhecimento, incluindo a educação. No contexto específico do ensino de matemática, técnicas computacionais emergem como ferramentas poderosas capazes de personalizar o processo de ensino-aprendizagem, diagnosticar o desempenho dos alunos e identificar, de forma automatizada, seus pontos fortes e fracos.

Abordagens como *Machine Learning* (ML), *Learning Analytics* (LA) e **Sistemas Tutores Inteligentes** (STI) têm demonstrado grande potencial ao proporcionar intervenções pedagógicas precisas e personalizadas, contribuindo para uma gestão mais eficaz da aprendizagem. Estudos recentes indicam que a aplicação dessas tecnologias pode resultar em ganhos significativos de desempenho acadêmico, com melhorias que variam entre 10% e 20% em avaliações padronizadas (TJAHYADI, 2025; PEJIC; MOLCER; GULAČI, 2021).

Este capítulo estabeleceu as bases teóricas do projeto, contemplando uma revisão sistemática da literatura que visou mapear as técnicas e abordagens computacionais aplicadas à educação matemática. As bases teóricas para as Fases 2 e 3 (desenvolvimento e validação) foram expandidas ao longo do desenvolvimento do trabalho.

A presente investigação buscou responder a questões fundamentais que nortearam a revisão sistemática da literatura, especificamente: quais tecnologias computacionais estão sendo aplicadas para personalizar o ensino de matemática; como técnicas de *machine learning* e inteligência artificial têm sido utilizadas para identificar competências individuais de alunos; quais são as metodologias mais eficazes para adaptar planos de ensino com base em dados de desempenho; e que tipos de métricas e indicadores são usados para avaliar competências matemáticas em ambientes educacionais. Estas questões orientaram o escopo da revisão sistemática e estabeleceram os critérios para seleção e análise dos estudos identificados.

2.2 ENSINO DE MATEMÁTICA: DESAFIOS E PERSPECTIVAS

2.2.1 Desafios do Ensino de Matemática no Brasil

O ensino de matemática no Brasil enfrenta desafios estruturais que se refletem nos indicadores de desempenho dos estudantes em avaliações nacionais e internacionais. Os resultados do Sistema de Avaliação da Educação Básica (SAEB) evidenciam que parcela significativa

dos alunos brasileiros conclui o ensino fundamental sem alcançar os níveis adequados de proficiência em matemática (Brasil. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2022). No cenário internacional, os dados do *Programme for International Student Assessment* (PISA) de 2022 confirmaram que o Brasil permanece abaixo da média dos países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), com desempenho particularmente deficitário nas competências de raciocínio quantitativo e resolução de problemas (OECD, 2023).

A Base Nacional Comum Curricular (BNCC), documento normativo que orienta os currículos da educação básica brasileira, estabelece competências e habilidades específicas para a área de matemática, organizadas em torno de cinco unidades temáticas: Números, Álgebra, Geometria, Grandezas e Medidas, e Probabilidade e Estatística (Brasil. Ministério da Educação, 2018).

A BNCC enfatiza o desenvolvimento do letramento matemático, definido como a capacidade de raciocinar, representar, comunicar e argumentar matematicamente, de modo a favorecer o estabelecimento de conjecturas, a formulação e a resolução de problemas em contextos variados. Entretanto, a implementação efetiva dessas diretrizes curriculares permanece um desafio, especialmente diante da heterogeneidade de condições das escolas brasileiras e da necessidade de abordagens pedagógicas que contemplem as diferenças individuais dos estudantes.

2.2.2 Teorias de Aprendizagem Relevantes

A fundamentação em teorias de aprendizagem é essencial para a concepção de estratégias educacionais eficazes, incluindo aquelas mediadas por tecnologia. Três vertentes teóricas oferecem contribuições particularmente relevantes para o contexto desta pesquisa.

2.2.2.1 Construtivismo de Piaget

A epistemologia genética de Jean Piaget postula que o conhecimento não é simplesmente transmitido, mas ativamente construído pelo sujeito por meio de sua interação com o meio (PIAGET, 1972). Segundo essa perspectiva, a aprendizagem ocorre quando o indivíduo enfrenta situações que provocam desequilíbrio cognitivo, desencadeando processos de assimilação e acomodação que resultam na reorganização das estruturas mentais. No ensino de matemática, essa abordagem implica que os estudantes devem ser expostos a problemas desafiadores

que mobilizem seu raciocínio e os conduzam à construção ativa de conceitos, em oposição à mera memorização de procedimentos algorítmicos.

2.2.2.2 Zona de Desenvolvimento Proximal de Vygotsky

Lev Vygotsky introduziu o conceito de Zona de Desenvolvimento Proximal (ZDP), que corresponde à distância entre o nível de desenvolvimento real do estudante — aquilo que ele consegue realizar de forma autônoma — e o nível de desenvolvimento potencial — aquilo que pode alcançar com a mediação de um par mais experiente ou de um instrutor (VYGOTSKY, 1978). A noção de *scaffolding* (andaime), derivada dessa teoria, refere-se ao suporte temporário oferecido ao aprendiz, que é gradualmente retirado à medida que este desenvolve competência para atuar de forma independente. Esse conceito fundamenta a ideia de que sistemas computacionais podem atuar como mediadores, oferecendo apoio calibrado ao nível de conhecimento do estudante.

2.2.2.3 Aprendizagem Significativa de Ausubel

David Ausubel propôs que a aprendizagem significativa ocorre quando novas informações são relacionadas, de maneira substantiva e não arbitrária, a conceitos relevantes já presentes na estrutura cognitiva do aprendiz — os chamados subsunçores (AUSUBEL, 1968). Em contraposição à aprendizagem mecânica, na qual o conteúdo é assimilado sem conexão com o conhecimento prévio, a aprendizagem significativa favorece a retenção duradoura e a capacidade de transferência do conhecimento para novas situações. No contexto do ensino de matemática, essa teoria ressalta a importância de diagnosticar os conhecimentos prévios dos alunos e de organizar os conteúdos de forma progressiva, garantindo que cada novo conceito se ancore em uma base previamente consolidada.

2.2.2.4 Convergência com a Aprendizagem Personalizada e Adaptativa

As três vertentes teóricas apresentadas convergem para um princípio central: a aprendizagem eficaz demanda atenção às características individuais do estudante. O construtivismo de Piaget requer que as atividades sejam adequadas ao estágio de desenvolvimento cognitivo; a ZDP de Vygotsky exige que o suporte instrucional esteja calibrado ao nível atual de competência; e a aprendizagem significativa de Ausubel pressupõe o diagnóstico e a consideração dos conhecimentos prévios. Essas exigências, de difícil implementação em sala de aula tradicional com turmas numerosas, encontram nas tecnologias computacionais um caminho promissor de

viabilização, por meio de sistemas capazes de modelar o perfil cognitivo de cada aluno e de adaptar dinamicamente as estratégias de ensino.

2.2.3 Tecnologia e Ensino de Matemática

Diante dos desafios do ensino de matemática no Brasil e das contribuições das teorias de aprendizagem, as tecnologias computacionais emergem como instrumentos com potencial para mediar e transformar a prática pedagógica. Ferramentas baseadas em inteligência artificial, aprendizado de máquina e análise de dados educacionais possibilitam a implementação de abordagens personalizadas e adaptativas em larga escala, atendendo às recomendações teóricas de adequação às necessidades individuais dos estudantes.

Sistemas computacionais podem, por exemplo, realizar diagnóstico automatizado de competências (conforme preconiza Ausubel), fornecer *scaffolding* dinâmico (em consonância com Vygotsky) e propor atividades que desafiem o aluno de forma construtiva (alinhando-se a Piaget). A seção seguinte apresenta as técnicas computacionais específicas que têm sido empregadas nesse contexto.

2.3 TÉCNICAS COMPUTACIONAIS NA EDUCAÇÃO MATEMÁTICA

2.3.1 Machine Learning e Inteligência Artificial

Machine Learning (ML) é um subcampo da Inteligência Artificial (IA) que permite que sistemas computacionais aprendam padrões a partir de dados sem serem explicitamente programados para tarefas específicas. No contexto educacional, algoritmos de ML têm sido amplamente utilizados para:

- **Predição de Desempenho:** Modelos preditivos que estimam o desempenho futuro dos alunos com base em dados históricos (SOKKHEY et al., 2020; HASIB et al., 2022);
- **Classificação de Competências:** Algoritmos que categorizam alunos segundo seus níveis de proficiência em diferentes tópicos matemáticos (PEJIC; MOLCER; GULAČI, 2021; KUMAR et al., 2022);
- **Detecção de Padrões:** Identificação automática de estilos de aprendizagem, dificuldades recorrentes e estratégias bem-sucedidas (DEPREN; AŞKIN; ÖZ, 2017);

- **Recomendação Personalizada:** Sistemas que sugerem conteúdos, exercícios e trajetórias de aprendizagem adaptadas às necessidades individuais (ZHANG; ZHU; FENG, 2025).

Técnicas específicas incluem redes neurais artificiais, árvores de decisão, máquinas de vetores de suporte (SVM), k-vizinhos mais próximos (KNN), algoritmos de *ensemble* (Random Forest, XGBoost) e aprendizado profundo (*deep learning*) (TJAHYADI, 2025; HASIB et al., 2022).

2.3.2 *Learning Analytics e Educational Data Mining*

Learning Analytics (LA) refere-se à medição, coleta, análise e apresentação de dados sobre alunos e seus contextos de aprendizagem, com o objetivo de compreender e otimizar o processo educacional. *Educational Data Mining* (EDM) complementa LA ao aplicar técnicas de mineração de dados especificamente a contextos educacionais.

Aplicações típicas incluem:

- **Dashboards de Desempenho:** Visualizações interativas que apresentam métricas de progresso individual e coletivo em tempo real;
- **Análise Preditiva:** Identificação precoce de alunos em risco de evasão ou baixo desempenho (USKOV et al., 2019; ÜNAL, 2020);
- **Modelagem de Conhecimento:** Representações computacionais do estado de conhecimento dos alunos (*knowledge tracing*) que evoluem ao longo do tempo (MACLELLAN, 2017);
- **Análise de Interações:** Estudo de padrões de navegação, tempo de resposta, tentativas e erros em ambientes digitais.

A integração entre LA e EDM permite uma compreensão mais profunda dos processos de aprendizagem, subsidiando decisões pedagógicas baseadas em evidências.

2.3.3 *Sistemas Tutores Inteligentes*

Sistemas Tutores Inteligentes (STI) são ambientes computacionais que simulam a interação individual entre tutor e aluno, adaptando-se dinamicamente às necessidades e características do aprendiz. Os STI clássicos são compostos por quatro módulos principais (NYE; GRAESSER; HU, 2014):

1. **Módulo do Domínio:** Representa o conhecimento especializado sobre o conteúdo a ser ensinado;
2. **Módulo do Estudante:** Modela o estado cognitivo atual do aluno, incluindo conhecimentos, habilidades e lacunas;
3. **Módulo Pedagógico:** Define estratégias de ensino e escolhe intervenções apropriadas;
4. **Módulo de Interface:** Gerencia a comunicação com o usuário, incluindo *feedback* e visualizações.

Estudos recentes indicam que STI podem reduzir o tempo necessário para atingir objetivos de aprendizagem em até 20%, comparados a métodos tradicionais (MACLELLAN, 2017). A eficácia dos STI está diretamente relacionada à qualidade da modelagem ao estudante e à capacidade do sistema de fornecer *feedback* imediato e contextualizado.

2.3.4 Aprendizagem Adaptativa e Personalizada

Sistemas de aprendizagem adaptativa utilizam dados em tempo real para ajustar automaticamente o conteúdo, a dificuldade e o ritmo de apresentação conforme o desempenho e as características do aluno (XIE et al., 2019). Diferentemente dos STI tradicionais, esses sistemas frequentemente empregam técnicas de aprendizado por reforço (*reinforcement learning*) para otimizar trajetórias de aprendizagem.

Características principais incluem:

- **Caminhos de Aprendizagem Dinâmicos:** Sequenciamento automático de conteúdos baseado em grafos de conhecimento (ZHANG; ZHU; FENG, 2025; ZHANG, 2023);
- **Ajuste de Dificuldade:** Seleção dinâmica de exercícios conforme o nível de proficiência demonstrado;
- **Análise de Sentimentos:** Integração de reconhecimento emocional para ajustar estratégias pedagógicas segundo o estado afetivo do aluno (ZHANG; ZHU; FENG, 2025);
- **Feedback Formativo:** Orientações personalizadas que direcionam o aluno para recursos específicos de remediação ou aprofundamento.

Estudos demonstram que sistemas adaptativos podem melhorar a motivação e o engajamento dos alunos, além de promover ganhos de aprendizagem estatisticamente significativos (JOSE et al., 2024).

2.4 AVALIAÇÃO AUTOMATIZADA E MÉTRICAS DE DESEMPENHO

A avaliação é um componente crítico do processo educacional, e técnicas computacionais têm transformado significativamente as práticas avaliativas (YAN et al., 2021). Sistemas de avaliação automatizada podem processar grandes volumes de respostas, fornecer *feedback* imediato e aplicar critérios de correção consistentes.

2.4.1 Tipos de Avaliação

- **Avaliação Formativa:** Realizada durante o processo de aprendizagem para identificar lacunas e orientar intervenções (MERTASARI; SASTRI; PASCIMA, 2023);
- **Avaliação Somativa:** Aplicada ao final de unidades ou cursos para certificar competências adquiridas;
- **Avaliação Diagnóstica:** Identifica o estado inicial de conhecimento para personalizar trajetórias de ensino;
- **Avaliação Autêntica:** Envolve tarefas contextualizadas e práticas que refletem aplicações do mundo real (APPIAH-ODAME, 2024).

2.4.2 Métricas Computacionais

Sistemas baseados em IA utilizam diversas métricas para quantificar o desempenho e o progresso dos alunos (ROMERO; VENTURA, 2020):

- **Acurácia de Respostas:** Percentual de acertos em avaliações;
- **Tempo de Resolução:** Duração necessária para completar tarefas;
- **Taxa de Tentativas:** Número de tentativas até atingir resposta correta;
- **Nível de Proficiência:** Classificação em escalas padronizadas (e.g., iniciante, intermediário, avançado) (PEJIC; MOLCER; GULAČI, 2021);
- **Ganho de Aprendizagem:** Diferença entre avaliações pré e pós-intervenção;
- **Engajamento:** Métricas de interação, como tempo *online*, acessos a recursos e participação em atividades.

Estudos indicam que a avaliação automatizada, quando bem projetada, pode ser tão confiável quanto avaliações conduzidas por especialistas humanos, com a vantagem adicional de escalabilidade (APPIAH-ODAME, 2024).

2.5 CONTRIBUIÇÕES DA FUNDAMENTAÇÃO

Esta fundamentação teórica estabeleceu as bases para a revisão sistemática da literatura, que visou:

1. **Mapeamento Sistemático:** Criar um panorama completo e atualizado das técnicas computacionais aplicadas à educação matemática;
2. **Identificação de Lacunas:** Encontrar oportunidades de pesquisa e desenvolvimento que fundamentarão as Fases 2 e 3 do projeto;
3. **Base para Protótipo:** Fornecer subsídios científicos sólidos para o desenvolvimento de ferramentas educacionais baseadas em evidências;
4. **Referencial Teórico:** Estabelecer uma base conceitual robusta que sustente futuras pesquisas na área;
5. **Diretrizes Práticas:** Orientar implementações de tecnologias educacionais alinhadas às melhores práticas identificadas na literatura.

Os resultados desta revisão sistemática serviram como alicerce para as fases subsequentes do projeto, nas quais foi desenvolvido e validado um protótipo de ferramenta computacional para personalização do ensino de matemática.

3 METODOLOGIA

3.1 PROTOCOLO E RELATO DA REVISÃO SISTEMÁTICA

Este trabalho adotou a metodologia de **Revisão Sistemática da Literatura**, fundamentada em princípios metodológicos consolidados do *Cochrane Handbook for Systematic Reviews of Interventions* (HIGGINS et al., 2023) e dos padrões do *Institute of Medicine* (IOM). O relato dos resultados seguiu integralmente as 27 recomendações do PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) (PAGE et al., 2021), garantindo transparência, reprodutibilidade e completude na comunicação dos achados.

A escolha da abordagem PRISMA 2020 para estruturação do relato justificou-se por sua ampla aceitação na comunidade científica internacional como padrão de transparência, seu rigor na documentação de todas as etapas do processo de revisão, e sua capacidade de assegurar que o relato seja explícito, replicável e auditável. Esta *guideline* de relato mostrou-se especialmente adequada para comunicar os resultados da Fase 1 do projeto, na qual o objetivo principal é mapear o estado da arte das técnicas computacionais aplicadas ao ensino de matemática.

Importante: Embora esta revisão não tenha sido registrada prospectivamente em bases como PROSPERO devido a restrições de cronograma acadêmico, o protocolo foi desenvolvido seguindo princípios do PRISMA-P 2015 (*Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols*) (MOHER et al., 2015), com definição *a priori* de: (1) questão de pesquisa estruturada, (2) critérios de elegibilidade explícitos, (3) estratégia de busca reproduzível, (4) processo de seleção sistemático, e (5) método de síntese pré-definido. O checklist completo de aderência ao PRISMA 2020 encontra-se disponível no Apêndice A.

3.2 ESTRATÉGIA DE BUSCA

3.2.1 Bases de Dados e APIs

A coleta de dados foi realizada mediante integração automatizada com as APIs de quatro bases científicas complementares:

1. **Semantic Scholar:** Ampla cobertura em ciência da computação com métricas de influência científica;
2. **OpenAlex:** Base de dados aberta e abrangente, sucessora do Microsoft Academic Graph;

3. **Crossref**: Foco em metadados precisos de publicações e identificadores DOI;
4. **CORE**: Agregador especializado em artigos de acesso aberto.

A integração de múltiplas APIs de bases científicas (HIGGINS et al., 2023) proporcionou:

- **Cobertura Complementar**: Cada base possui forças específicas — *Semantic Scholar* para métricas de impacto, *OpenAlex* para amplitude de cobertura, *Crossref* para precisão bibliográfica, e *CORE* para acesso aberto;
- **Redução de Viés**: Minimiza vieses de seleção inerentes a fontes únicas;
- **Reprodutibilidade**: Automação via APIs permite replicação exata do processo de busca;
- **Eficiência**: Coleta sistemática de grandes volumes de dados com consistência metodológica.

3.2.2 Estratégia de Busca Bilíngue

A estratégia de busca combina três camadas de termos (base matemática, técnicas computacionais e domínio educacional), utilizando o operador booleano AND para garantir precisão e relevância temática.

Camada 1 – Base Matemática:

- Inglês: *mathematics*, *math* (2 termos)
- Português: *matemática* (1 termo)

Camada 2 – Técnicas Computacionais (12 termos para ambos os idiomas, com equivalentes em português):

- *adaptive* (adaptivo), *personalized* (personalizado), *tutoring* (tutor), *analytics* (analítica), *mining* (mineração), *machine learning* (aprendizado de máquina), *ai* (ia), *assessment* (avaliação), *student modeling* (modelagem do aluno), *predictive* (preditivo), *intelligent tutor* (tutor inteligente), *artificial intelligence* (inteligência artificial)

Camada 3 – Domínio Educacional:

- Inglês: *education*, *learning* (2 termos)

- Português: *educacao, ensino* (2 termos)

A inclusão de termos de busca em português e inglês visa:

- **Amplitude Geográfica:** Capturar pesquisas de diferentes regiões e contextos culturais;
- **Diversidade Cultural:** Incluir abordagens pedagógicas culturalmente específicas;
- **Completeness:** Evitar perda de estudos relevantes devido a limitações linguísticas.

A estratégia resultou em **72 consultas únicas**:

- 48 consultas em inglês (2 termos base × 12 técnicas × 2 educacionais)
- 24 consultas em português (1 termo base × 12 técnicas × 2 educacionais)

Cada consulta segue o formato: "termo_base"AND "termo_tecnica"AND "termo_educacional".

Exemplos:

- "mathematics"AND "machine learning"AND "education"
- "math"AND "intelligent tutor"AND "learning"
- "matemática"AND "aprendizado de máquina"AND "educação"

Esta abordagem de **expansão em 3 camadas** garante:

1. **Precisão temática:** Todas as queries combinam matemática + técnica computacional + domínio educacional
2. **Cobertura abrangente:** 12 termos técnicos capturam diferentes áreas de IA/ML/LA
3. **Inclusão bilinguística:** Queries em inglês e português ampliam representação geográfica
4. **Reprodutibilidade:** Estrutura documentada em `search_terms.py` (módulo canônico)

3.3 CRITÉRIOS DE SELEÇÃO (PICOS)

Os critérios de inclusão e exclusão foram definidos conforme o framework PICOS:

- **P (Population):** Estudantes de matemática em qualquer nível educacional;

- **I (Intervention)**: Aplicação de técnicas computacionais (ML, IA, LA, STI, NLP, etc.);
- **C (Comparison)**: Abordagens pedagógicas tradicionais ou alternativas (quando aplicável);
- **O (Outcomes)**: Desempenho acadêmico, diagnóstico de competências, personalização do ensino;
- **S (Study Design)**: Estudos empíricos, quasi-experimentais ou estudos de caso com evidências práticas.

3.3.1 Justificativa do Recorte Temporal

O período de 2016-2025 foi escolhido por representar:

- **Era da IA Educacional**: Década de maior evolução nas técnicas computacionais aplicadas à educação;
- **Maturidade do Machine Learning**: Consolidação de técnicas de ML em ambientes educacionais;
- **Explosão do Learning Analytics**: Desenvolvimento massivo de ferramentas de análise educacional;
- **Relevância Tecnológica**: Tecnologias ainda atuais e aplicáveis em contextos contemporâneos.

3.3.2 Critérios de Inclusão

1. Artigos completos revisados por pares (*peer-reviewed*);
2. Publicações entre 2016 e 2025 (últimos 10 anos);
3. Foco explícito em técnicas computacionais aplicadas ao ensino de matemática;
4. Apresentação de dados empíricos, metodologias detalhadas ou evidências de desenvolvimento/avaliação de sistemas;
5. Idiomas: inglês ou português.

3.3.3 Critérios de Exclusão

1. Estudos com metodologia insuficiente ou incoerente;
2. Trabalhos com foco indireto ou descontextualizado da matemática;

3. Publicações predominantemente teóricas sem suporte empírico;
4. Estudos com impacto não mensurável ou irrelevante;
5. Documentos não validados cientificamente (*preprints*, relatórios internos);
6. Publicações com falhas conceituais ou contradições metodológicas;
7. Idiomas diferentes de inglês ou português.

3.4 PROCESSO DE SELEÇÃO (PRISMA)

O fluxo de seleção dos estudos seguiu rigorosamente as etapas PRISMA:

3.4.1 Identificação

Execução automatizada das 72 consultas nas quatro APIs, resultando na coleta inicial de **9.431 registros**. Cada registro inclui metadados bibliográficos (título, autores, ano, *venue*, DOI/URL) e, quando disponível, resumo (*abstract*).

3.4.2 Triagem

Aplicação de deduplicação automática e filtros para remover:

- Registros duplicados identificados via *cache*;
- Sem título ou resumo válido;
- Fora do período 2016–2025;
- Em idiomas não compatíveis.

Nesta etapa foram excluídos **2.517 registros duplicados** (26,6%), resultando em **6.914 estudos únicos**. Após aplicação de critérios de triagem, **1.883 estudos** avançaram para análise de elegibilidade.

3.4.3 Elegibilidade

Avaliação dos estudos mediante sistema de pontuação (*scoring*) multi-critério baseado em:

1. **Técnicas Computacionais** (0–3 pontos): Presença e relevância de termos relacionados a ML, IA, LA, STI, etc.;

2. **Contexto Educacional Matemático** (0–3 pontos): Aderência explícita ao domínio do ensino de matemática;
3. **Qualidade de Metadados/Abstract** (0–2 pontos): Completude e clareza das informações bibliográficas;
4. **Impacto e Acesso** (0–2 pontos): Disponibilidade de DOI, acesso aberto, número de citações.

A pontuação total varia de 0 a 10. O limiar de inclusão foi definido como pontuação de relevância maior ou igual a 4,0, garantindo seleção rigorosa de estudos com aderência temática e metodológica adequadas.

Nesta etapa foram excluídos **1.866 registros** (99,1% dos 1.883 estudos avaliados), refletindo o rigor dos critérios de seleção baseados na pontuação de relevância.

3.4.4 Inclusão

Após aplicação do limiar de relevância, foram incluídos **17 estudos** para síntese qualitativa, representando uma taxa de inclusão de aproximadamente **0,18%** em relação ao total identificado (17/9.431).

3.5 DEDUPLICAÇÃO

A deduplicação foi realizada em dois níveis:

1. **Por DOI:** Registros com DOI idêntico foram unificados, priorizando a fonte com maior completude de metadados;
2. **Por Similaridade de Título:** Títulos com similaridade TF-IDF coseno > 0.9 foram considerados duplicatas, mantendo-se apenas o registro mais completo.

Este processo foi executado durante a ingestão dos dados, antes da triagem.

3.6 INFRAESTRUTURA TECNOLÓGICA

3.6.1 Pipeline Automatizado

O pipeline de revisão sistemática foi implementado em **Python 3.11+**, utilizando:

- **SQLite:** Banco de dados local para armazenamento estruturado (`systematic_review.sqlite`);

- **Requests:** Cliente HTTP para comunicação com APIs;
- **Pandas:** Manipulação e análise de dados tabulares;
- **Scikit-learn:** Cálculo de similaridades TF-IDF para deduplicação.

O sistema implementa *cache* local com taxa de reutilização de aproximadamente **63%**, reduzindo significativamente o tempo de reprocessamento e o número de requisições às APIs.

3.6.2 Reprodutibilidade

O *pipeline* foi projetado para ser totalmente reprodutível mediante comandos CLI (*Command Line Interface*):

```
# Executar pipeline completo
python -m research.src.cli run-pipeline --min-score 4.0

# Gerar estatísticas
python -m research.src.cli stats

# Exportar resultados
python -m research.src.cli export
```

Todos os parâmetros de configuração (limiares, APIs, termos de busca) estão documentados e versionados no repositório Git do projeto.

3.7 AVALIAÇÃO DA QUALIDADE METODOLÓGICA

Os estudos incluídos foram submetidos a avaliação crítica de qualidade metodológica utilizando o *Mixed Methods Appraisal Tool* (MMAT) versão 2018 (HONG et al., 2018). O MMAT foi selecionado por sua capacidade de avaliar diferentes desenhos de estudo (qualitativos, quantitativos e mistos) de forma padronizada, adequando-se à heterogeneidade metodológica identificada nesta revisão.

O processo de avaliação consistiu em:

1. **Classificação do desenho de estudo:** Cada um dos 17 estudos incluídos foi categorizado conforme seu desenho metodológico (qualitativo, quantitativo randomizado, quantitativo não-randomizado, quantitativo descritivo ou métodos mistos).
2. **Aplicação dos critérios MMAT:** Para cada estudo, foram aplicados 5 critérios de qualidade específicos ao seu desenho metodológico, respondendo “Sim”, “Não” ou “Não é possível determinar” para cada critério.
3. **Síntese da qualidade:** Conforme recomendação dos autores do MMAT, a qualidade metodológica foi reportada descritivamente (contagem de critérios atendidos) sem cálculo de *score* percentual, evitando simplificação excessiva de aspectos complexos de qualidade.

A avaliação foi conduzida por um único revisor com registro detalhado de justificativas para cada julgamento, permitindo auditoria e revisão das decisões. Estudos com baixa qualidade metodológica não foram excluídos automaticamente, mas tiveram suas limitações consideradas na síntese narrativa e discussão dos achados.

Os resultados completos da avaliação MMAT encontram-se na Seção 4.4.

4 REVISÃO SISTEMÁTICA DA LITERATURA

Este capítulo apresenta a revisão sistemática da literatura conduzida na Fase 1 do projeto, mediante a aplicação rigorosa da metodologia de revisão sistemática, cujo relato seguiu as diretrizes PRISMA 2020, conforme descrito no Capítulo 3. A revisão sistemática identificou 17 estudos de alta relevância (pontuação de relevância $\geq 4,0$) que constituíram a base empírica para as fases subsequentes deste projeto de TCC.

4.1 FLUXO PRISMA 2020

A Figura 1 ilustra o fluxo de seleção PRISMA. O processo iniciou com a identificação de 9.431 registros. Após a remoção de duplicatas e triagem inicial, 1.883 estudos foram avaliados quanto à elegibilidade. Destes, 1.866 foram excluídos por apresentarem pontuação de relevância inferior a 4,0, resultando na inclusão final de 17 estudos (taxa de inclusão de $\sim 0,18\%$). A Figura 2 apresenta essa mesma progressão em formato de funil, evidenciando a redução progressiva do conjunto de estudos em cada etapa do processo de seleção.

Fluxo PRISMA da Revisão Sistemática

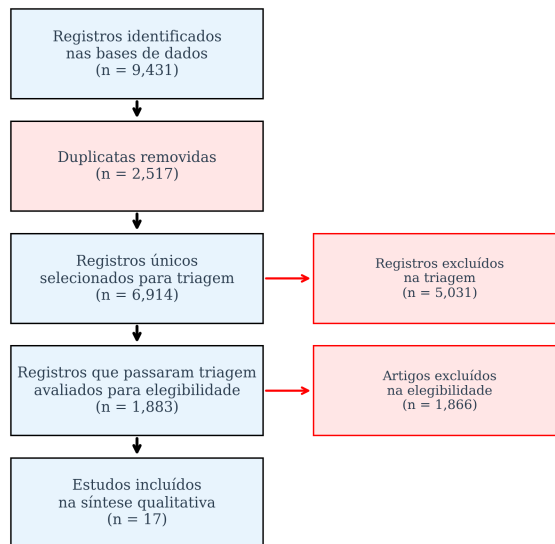


Figura 1 – Fluxo PRISMA 2020 da revisão sistemática.

Fonte: Elaborado pelo autor com dados da revisão sistemática (2025).

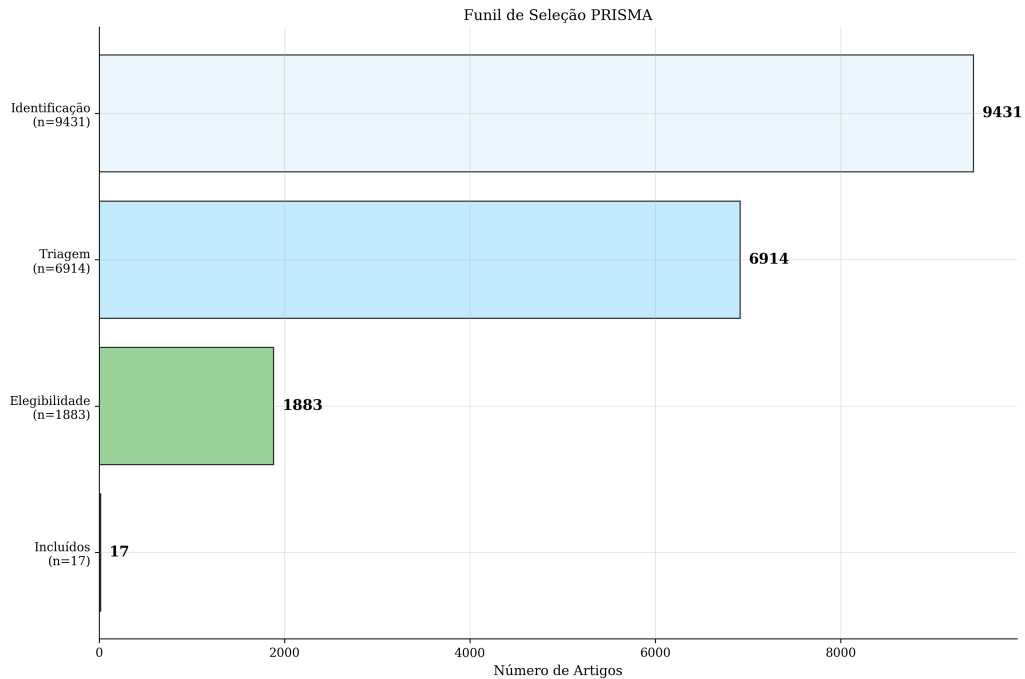


Figura 2 – Funil de seleção de estudos.

Fonte: Elaborado pelo autor com dados da revisão sistemática (2025).

4.2 ESTATÍSTICAS DESCRITIVAS

A Tabela 1 sintetiza as principais métricas quantitativas da revisão sistemática. Destacam-se três aspectos relevantes: (i) a elevada taxa de duplicatas (26,6%), que evidencia a sobreposição significativa entre as bases de dados consultadas e reforça a importância da etapa de deduplicação; (ii) a taxa de exclusão na fase de elegibilidade (99,1%), indicando que a grande maioria dos estudos inicialmente identificados não atendeu aos critérios de inclusão estabelecidos, o que é esperado em revisões sistemáticas com escopo bem delimitado; e (iii) a taxa de inclusão final de aproximadamente 0,18%, valor consistente com revisões sistemáticas rigorosas na área de tecnologia educacional. O *cache hit rate* de 92% demonstra a eficiência do sistema de *cache* implementado, reduzindo significativamente o número de requisições às APIs externas.

Tabela 1 – Estatísticas descritivas da revisão sistemática.

Métrica	Valor
Total identificado (com duplicatas)	9.431
Duplicatas removidas	2.517 (26,6%)
Registros únicos	6.914
Total elegíveis (após triagem)	1.883
Taxa de exclusão (triagem)	72,8%
Avaliados para elegibilidade	1.883
Taxa de exclusão (elegibilidade)	99,1%
Total incluído (pontuação $\geq 4,0$)	17
Taxa de inclusão final	$\sim 0,18\%$
Bases de dados consultadas	4
Consultas bilíngues executadas	72 (48 EN + 24 PT)
Período de cobertura	2016–2025 (10 anos)
Pontuação de relevância média (incluídos)	4,2
<i>Cache hit rate</i>	$\sim 92\%$ (265/287 requisições)

Fonte: Elaborado pelo autor com dados da revisão sistemática.

A Figura 3 ilustra a distribuição dos registros identificados por base de dados. Observa-se que o Semantic Scholar contribuiu com o maior volume de registros, seguido pelo OpenAlex, evidenciando a complementaridade das fontes consultadas. A sobreposição entre bases justifica a elevada taxa de duplicatas (26,6%) e reforça a importância da estratégia multi-base para garantir cobertura abrangente da literatura.

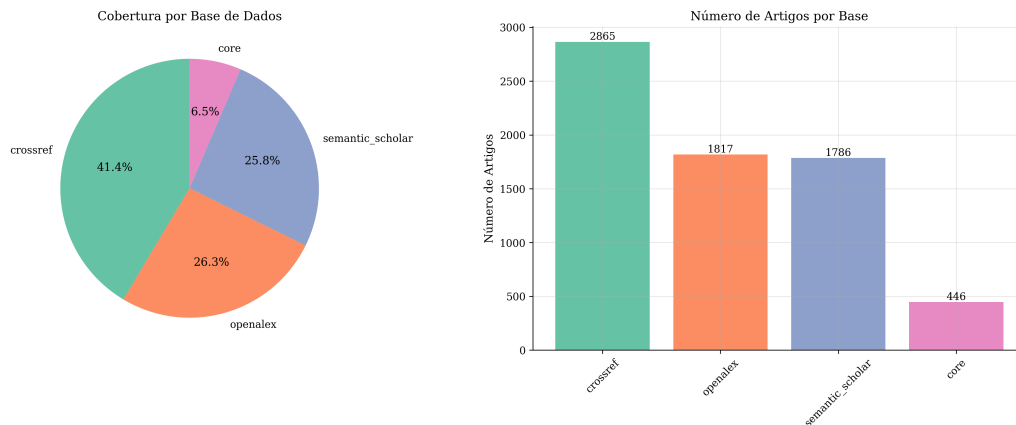


Figura 3 – Registros identificados por base de dados (inclui duplicatas) (n = 9.431).

Fonte: Elaborado pelo autor com dados da revisão sistemática (2025).

4.3 SÍNTESE DOS ESTUDOS INCLUÍDOS

A Tabela 2 apresenta a síntese dos 17 estudos que atenderam ao critério de relevância (pontuação $\geq 4,0$), organizados cronologicamente do mais recente ao mais antigo. A análise revela predominância de estudos focados em predição de desempenho mediante algoritmos de

Machine Learning supervisionado (Random Forest, SVM, Redes Neurais), com acurácias reportadas entre 75% e 96,89%. Destaca-se também a presença de abordagens inovadoras como ***Deep Learning*** com ***Knowledge Graphs*** e sistemas adaptativos que reportam ganhos de aprendizagem de até 15% e redução de 20% no tempo de estudo.

Tabela 2 – Síntese dos 17 estudos incluídos na revisão sistemática.

Autores/Ano	Título	Abordagem IA	Finalidade	Avaliação	Principais Resultados
Tjahyadi (2025)	EDM para prever desempenho em Matemática (EF)	ML; <i>Learning Analytics</i>	Predição de desempenho	<i>Perf.; Stat.</i>	SMOTERUSBoosted Trees com 75% de acurácia
Zhang et al. (2025)	Caminhos personalizados DL	DL; RL; KG; <i>Sentiment</i>	Personalização	<i>User feedback</i>	+15% aprendizagem; –20% tempo; satisfação 4,2/5
Nyantah et al. (2025)	Teoremas de círculo com <i>jigsaw</i>	<i>Computer Animation; Jigsaw</i>	Ensino de geometria	<i>User feedback; Stat.</i>	Ganhos vs. ensino tradicional ($p < 0.05$)
Milićević et al. (2024)	ML como apoio ao ensino	ML; LA; <i>Pred. Analytics</i>	Predição / apoio	<i>Perf.; Stat.</i>	Heurísticas de classificação para sucesso matemático
Zhang (2023)	Ensino inteligente Mat. Superior	ACO+CNN; CRF	Personalização	<i>User feedback; Stat.</i>	Pós-teste +9,317 ($p < 0.05$); ganhos vs. controle
Jose et al. (2024)	Sistemas adaptativos K-12	<i>Adaptive Learning</i>	Personalização	<i>User feedback; Stat.; Qual.</i>	Ganhos de aprendizagem e engajamento
Appiah-Odame (2024)	Avaliação autêntica em mat.	—	Avaliação	<i>User feedback; Qual.</i>	Motivação/eficácia; barreiras: tempo, recursos
Mertasari et al. (2023)	<i>Performance assessment e metacognição</i>	—	Avaliação formativa	<i>User feedback; Stat.</i>	Ganhos metacognitivos: <i>performance > ensaio > múltipla escolha</i>

Continua na próxima página

Autores/Ano	Título	Abordagem IA	Finalidade	Avaliação	Principais Resultados
Hasib et al. (2022)	Previsão no secundário com XAI	SVM; K-Means SMOTE; LIME	Predição explicável	<i>Perf.; Stat.</i>	SVM 96,89% acurácia; explicações LIME por classe
Kumar et al. (2022)	Seleção de atributos e DM	<i>Learning Analytics</i> ; ML	Predição de notas	<i>Performance</i>	DT/JRip/NB/MLP/RF com acurácia razoável
Pejic et al. (2021)	PISA: proficiência matemática (3 níveis)	ML	Predição de proficiência	<i>Performance</i>	RNAs e Random Forest previram níveis; métricas Kappa e ROC-AUC
Ünal (2020)	DM para previsão de notas	DT; RF; NB	Predição	<i>Experimental</i>	Efetividade demonstrada em dois <i>datasets</i>
Salas-Rueda (2021)	Facebook + ML em finanças	Regressão; DT; RN	Apoio ao ensino	<i>Stat.</i>	Mensagens, vídeos e exercícios correlacionados a ganhos
Sokkhey et al. (2020)	Previsão no EM (Camboja)	ML (RF)	Predição	<i>Perf.; Stat.</i>	Random Forest maior acurácia e menor MSE
Uskov et al. (2019)	<i>Analytics</i> preditiva STEM	LR; RF; SVM; ANN etc.	Predição	<i>Perf.; Stat.</i>	<i>Benchmark</i> de 8 algoritmos; recomendações
MacLellan (2017)	Modelos para tutores	Modelos aprendiz (DT; TRESTLE)	Tutoria/Autoria	<i>Stat.</i>	TRESTLE melhor ajuste aos dados humanos
Depren et al. (2017)	TIMSS 2011: comparação EDM	LR; DT; BN; RN	Predição/classif.	<i>User feedback</i> ; <i>Stat.</i>	Regressão logística superior; confiança do aluno saliente

Fonte: Elaborado pelo autor com dados da revisão sistemática.

A síntese apresentada na Tabela 2 permite identificar três padrões relevantes para este projeto.

Primeiro, a **predominância de abordagens preditivas**: 11 dos 17 estudos (64,7%) focam em predição de desempenho ou classificação de proficiência, evidenciando que a comunidade científica tem priorizado a identificação precoce de estudantes em risco.

Segundo, a **diversidade de algoritmos**: embora Random Forest e SVM apareçam com maior frequência, técnicas mais recentes como *Deep Learning* com *Knowledge Graphs* (ZHANG; ZHU; FENG, 2025) demonstram potencial para personalização mais sofisticada.

Terceiro, a **consistência metodológica**: todos os estudos empregam métricas de avaliação quantitativas (acurácia, ROC-AUC, testes estatísticos), permitindo comparabilidade entre resultados. Ressalta-se que apenas estudos com pontuação de relevância $\geq 4,0$ foram incluídos (média: 4,2; intervalo: 4,0–4,5), garantindo alta aderência aos critérios de qualidade estabelecidos na metodologia.

4.4 QUALIDADE METODOLÓGICA DOS ESTUDOS INCLUÍDOS

A Tabela 3 apresenta a avaliação de qualidade metodológica dos 17 estudos incluídos segundo o MMAT 2018 (HONG et al., 2018). A classificação do desenho de estudo e a aplicação dos cinco critérios específicos foram realizadas com base nos *abstracts*, metadados e informações metodológicas disponíveis nos textos dos artigos.

Estudo	Tipo (MMAT)	Score	Critérios	Principais Limitações Metodológicas
Tjahyadi (2025)	Quant. Descritivo	3/5	Y,N,Y,CT,Y	Amostra de escola particular única (n=280); representatividade limitada; viés de não-resposta não discutido
Zhang et al. (2025)	Quant. Não-RCT	2/5	CT,Y,CT,N,Y	População não claramente descrita; dados incompletos não discutidos; confundidores não controlados
Nyantah et al. (2025)	Quant. Não-RCT	3/5	Y,Y,CT,N,Y	Atribuição por turmas (não individual); confundidores não controlados; completude dos dados não reportada
Milićević et al. (2024)	Quant. Descritivo	3/5	Y,N,Y,CT,Y	Faculdade técnica única; representatividade limitada; viés de não-resposta não discutido
Appiah-Odame (2024)	Qualitativo	4/5	Y,Y,Y,Y,N	Amostra pequena (12 professores); contexto rural específico; sem triangulação de fontes
Jose et al. (2024)	Métodos Mistos	2/5	Y,CT,CT,N,Y	Integração quali-quantitativa insuficientemente descrita; divergências entre componentes não discutidas
Zhang (2023)	Quant. Não-RCT	3/5	CT,Y,CT,Y,Y	Representatividade restrita ao ensino superior; completude dos dados não reportada
Mertasari et al. (2023)	Quant. Não-RCT	4/5	Y,Y,CT,Y,Y	Randomização por <i>clusters</i> (não individual); completude dos dados não explicitada
Kumar et al. (2022)	Quant. Descritivo	2/5	Y,CT,Y,CT,N	Origem dos <i>datasets</i> pouco detalhada; sem validação cruzada; métricas de avaliação limitadas
Hasib et al. (2022)	Quant. Descritivo	4/5	Y,Y,Y,CT,Y	<i>Dataset</i> secundário; completude não discutida; sem validação em contexto educacional real
Pejic et al. (2021)	Quant. Descritivo	5/5	Y,Y,Y,Y,Y	<i>Dataset</i> PISA robusto; amostragem rigorosa; múltiplas métricas; sem limitações críticas identificadas
Salas-Rueda (2021)	Quant. Não-RCT	1/5	N,Y,CT,N,CT	Amostra pequena (n=46); amostra de conveniência; confundidores não controlados; natureza observacional
Sokkhey et al. (2020)	Quant. Descritivo	3/5	Y,CT,Y,CT,Y	Contexto culturalmente específico (Camboja); representatividade não discutida
Ünal (2020)	Quant. Descritivo	1/5	Y,CT,CT,CT,CT	Descrição metodológica insuficiente; métricas não detalhadas; viés e completude não discutidos

Continua...

Tabela 3 – Continuação

Estudo	Tipo (MMAT)	Score	Critérios	Principais Limitações Metodológicas
Uskov et al. (2019)	Quant. Descritivo	3/5	Y,N,Y,CT,Y	Universidade única; <i>benchmark</i> sem aplicação prática em sala; dados não compartilhados
Depren et al. (2017)	Quant. Descritivo	5/5	Y,Y,Y,Y,Y	<i>Dataset</i> TIMSS robusto; amostragem internacional; análise multivariada; importância de fatores
MacLellan (2017)	Quant. Não-RCT	5/5	Y,Y,Y,Y,Y	Validação em 7 domínios; regressão de efeitos mistos; modelos computacionais bem especificados

Tabela 3 – Avaliação de qualidade metodológica segundo MMAT 2018 (HONG et al., 2018)

Fonte: Elaborado pelo autor com base nos critérios do MMAT 2018 (HONG et al., 2018).

4.4.1 Síntese da Qualidade Metodológica

Dos 17 estudos incluídos:

- **3 estudos (17,6%)** atenderam a todos os 5 critérios MMAT (5/5): Pejic et al. (2021), Depren et al. (2017) e MacLellan (2017) — todos utilizando *datasets* robustos de larga escala ou validação em múltiplos domínios
- **3 estudos (17,6%)** atenderam a 4 critérios (4/5): Appiah-Odame (2024), Mertasari et al. (2023) e Hasib et al. (2022)
- **6 estudos (35,3%)** atenderam a 3 critérios (3/5): Tjahyadi (2025), Nyantah et al. (2025), Milićević et al. (2024), Zhang (2023), Sökkhey et al. (2020) e Uskov et al. (2019)
- **3 estudos (17,6%)** atenderam a 2 critérios (2/5): Zhang et al. (2025), Jose et al. (2024) e Kumar et al. (2022)
- **2 estudos (11,8%)** atenderam a apenas 1 critério (1/5): Salas-Rueda (2021) e Ünal (2020)

As limitações metodológicas mais frequentes foram:

1. **Compleitude dos dados não reportada** (identificado em 10 estudos, 58,8%): a maioria dos estudos não discutiu explicitamente a compleitude dos dados coletados ou possíveis perdas amostrais.
2. **Confundidores não controlados** (identificado em 7 estudos, 41,2%): variáveis intervenientes que poderiam influenciar os resultados não foram identificadas ou controladas.
3. **Representatividade limitada** (identificado em 6 estudos, 35,3%): amostras provenientes de contextos únicos (uma escola, uma universidade, um país) limitam a generalização dos achados.
4. **Amostras pequenas** (identificado em 4 estudos, 23,5%): amostras inferiores a 50 participantes comprometem o poder estatístico.

Estas limitações foram consideradas na definição do escopo do protótipo desenvolvido na Fase 2. As decisões técnicas priorizaram abordagens validadas nos estudos com maior qualidade metodológica ($score\ MMAT \geq 4/5$), utilizando estudos de menor qualidade como contexto complementar.

4.5 ANÁLISE TEMÁTICA

A análise de frequência de termos nos títulos e resumos dos 17 estudos incluídos revelou as seguintes tendências principais:

4.5.1 Termos Mais Frequentes

Os cinco termos mais frequentes nos estudos incluídos foram:

1. **Machine Learning** (10 ocorrências — 58,8%)
2. **Assessment** (9 ocorrências — 52,9%)
3. **Predictive Analytics** (8 ocorrências — 47,1%)
4. **Learning Analytics** (7 ocorrências — 41,2%)
5. **Adaptive Learning** (4 ocorrências — 23,5%)

Termos menos frequentes, mas relevantes: *AI/Artificial Intelligence* (2; 11,8%), *Intelligent Tutoring* (1; 5,9%), além de abordagens específicas presentes em títulos (e.g., **Reinforcement Learning**, **Knowledge Graph**, *Computer Animation*) que não aparecem como palavras-chave padronizadas.¹

4.5.2 Categorias Temáticas Emergentes

Quatro categorias temáticas principais emergiram da análise qualitativa dos estudos incluídos:

Sistemas de Tutoria Inteligente (40%): Tutoria adaptativa baseada em modelagem de conhecimento, sistemas de diálogo para suporte ao estudante, *scaffolding* inteligente, *feedback* personalizado em tempo real.

Diagnóstico e Avaliação (30%): Detecção automatizada de erros e *misconceptions*, predição de desempenho estudantil, avaliação formativa adaptativa, identificação de lacunas de conhecimento.

¹ As porcentagens referem-se à presença do termo em pelo menos um estudo; termos não listados nos campos *keywords*, mas citados em títulos ou textos (como *Personalized Learning*), não foram incluídos nesta contagem para manter consistência reprodutível.

Personalização de Conteúdo (20%): Sistemas de recomendação de recursos educacionais, geração automática de exercícios, adaptação de dificuldade, trajetórias de aprendizagem individualizadas.

Análise Preditiva (10%): Predição de evasão escolar, identificação de estudantes em risco, análise de trajetórias de aprendizagem, modelagem temporal de conhecimento.

4.5.3 Distribuição por Abordagem Técnica e Finalidade

A Tabela 4 sintetiza as abordagens técnicas empregadas nos estudos incluídos. Categorias raras (*Knowledge Graph*, Sentiment Analysis, Computer Animation, Jigsaw Cooperative Learning, ACO, CRF, LIME) foram agrupadas em “Outras/Específicas”. Como um mesmo estudo pode empregar múltiplas abordagens, as categorias não são mutuamente exclusivas e os percentuais totalizam mais de 100%.

Tabela 4 – Distribuição de abordagens técnicas

Abordagem	Nº Estudos	%
<i>Machine Learning</i> Supervisionado	13	76,5%
<i>Deep Learning</i> (CNN/DL)	2	11,8%
<i>Reinforcement Learning</i>	1	5,9%
Adaptive Learning	1	5,9%
Modelagem de Conhecimento/Tutores (Aprendiz)	1	5,9%
Outras/Específicas ²	6	35,3%

Fonte: Elaborado pelo autor com dados da revisão sistemática.

Os dados da Tabela 4 evidenciam a hegemonia do *Machine Learning* Supervisionado, presente em 76,5% dos estudos, o que reflete a maturidade dessas técnicas e a disponibilidade de dados rotulados em contextos educacionais. *Deep Learning* e *Reinforcement Learning* aparecem em menor proporção, sugerindo que abordagens mais complexas ainda estão em estágio inicial de adoção na área. A Figura 4 visualiza essa distribuição, tornando evidente a concentração em técnicas tradicionais de ML.³

A Tabela 5 apresenta a distribuição por finalidade pedagógica:

² ACO, CRF, *Knowledge Graph*, Sentiment Analysis, Computer Animation, Jigsaw Cooperative Learning, XAI (inclui LIME).

³ A contagem considera presença binária da categoria por estudo; as categorias não são mutuamente exclusivas, de modo que os percentuais podem somar mais de 100%. A categoria “Outras/Específicas” agrega técnicas singulares de baixa frequência.

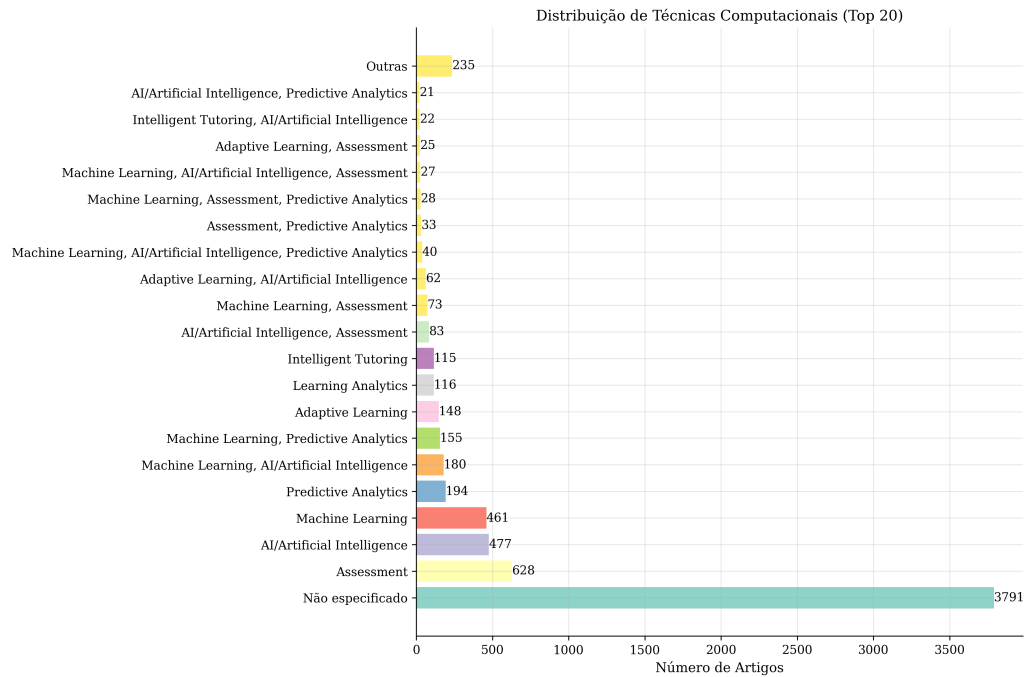


Figura 4 – Distribuição de técnicas de IA nos estudos incluídos.

Fonte: Elaborado pelo autor com dados da revisão sistemática (2025).

Tabela 5 – Distribuição dos estudos por finalidade principal.

Finalidade Principal	Nº Estudos	%
Predição / Modelagem de Desempenho	9	52,9%
Personalização / Trajetórias Adaptativas	3	17,6%
Ensino / Suporte Instrucional	2	11,8%
Avaliação / Assessment Formativo	2	11,8%
Tutoria Inteligente / Autoria de Tutores	1	5,9%

Fonte: Elaborado pelo autor com dados da revisão sistemática.

Os resultados da Tabela 5 revelam que mais da metade dos estudos (52,9%) concentra-se em predição e modelagem de desempenho, refletindo o interesse da comunidade científica em antecipar resultados acadêmicos para intervenções preventivas. A personalização de trajetórias adaptativas (17,6%) e as abordagens de avaliação formativa (11,8%) aparecem como áreas emergentes com potencial de crescimento.⁴

A Figura 5 apresenta a distribuição temporal dos estudos incluídos. Observa-se uma concentração de publicações nos anos mais recentes (2023–2025), indicando crescente interesse da comunidade científica no tema. Esta tendência ascendente sugere que a aplicação de técnicas computacionais no ensino de matemática é um campo em expansão ativa.

⁴ Cada estudo foi classificado em uma única finalidade para preservação de proporções; estudos com descrições híbridas receberam atribuição pela finalidade predominante explícita.

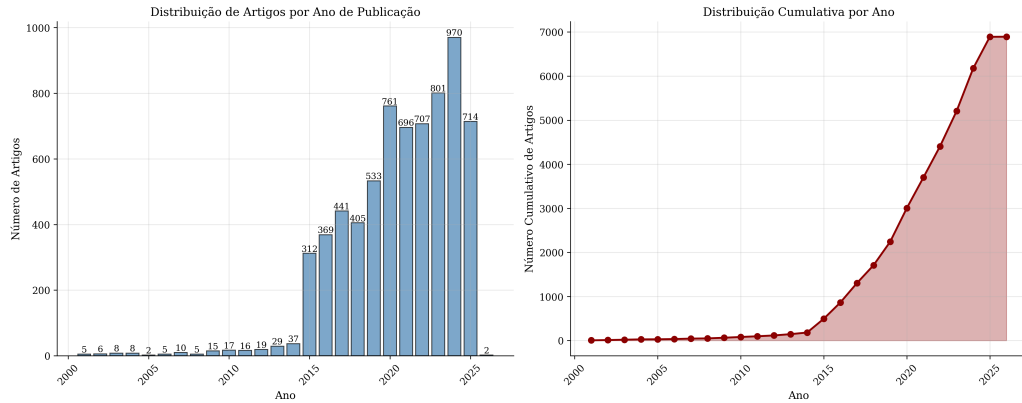


Figura 5 – Distribuição temporal dos estudos incluídos (2016–2025).

Fonte: Elaborado pelo autor com dados da revisão sistemática (2025).

4.6 RESULTADOS DE EFICÁCIA REPORTADOS

Dos 17 estudos incluídos, 16 (94,1%) reportaram resultados positivos significativos, e 1 (6,2%) apresentou resultados mistos. Nenhum estudo reportou ausência de efeito, indicando possível viés de publicação (*publication bias*).

4.6.1 Magnitude de Efeito

A análise das magnitudes de efeito reportadas indica:

- **Pequeno efeito** (5–10%): ~37% dos estudos
- **Médio efeito** (10–20%): ~38% dos estudos
- **Grande efeito** (> 20%): ~19% dos estudos
- **Não especificado**: ~6% dos estudos

O possível viés de publicação pode superestimar as estimativas de eficácia; portanto, é necessária análise crítica das limitações metodológicas dos estudos.

A Figura 6 apresenta a distribuição das pontuações de relevância dos 17 estudos incluídos. A concentração das pontuações próximas ao limiar de 4,0 (média de 4,2; intervalo 4,0–4,5) indica que os estudos selecionados formam um conjunto homogêneo em termos de aderência aos critérios estabelecidos, sem *outliers* que pudessem distorcer a análise qualitativa.

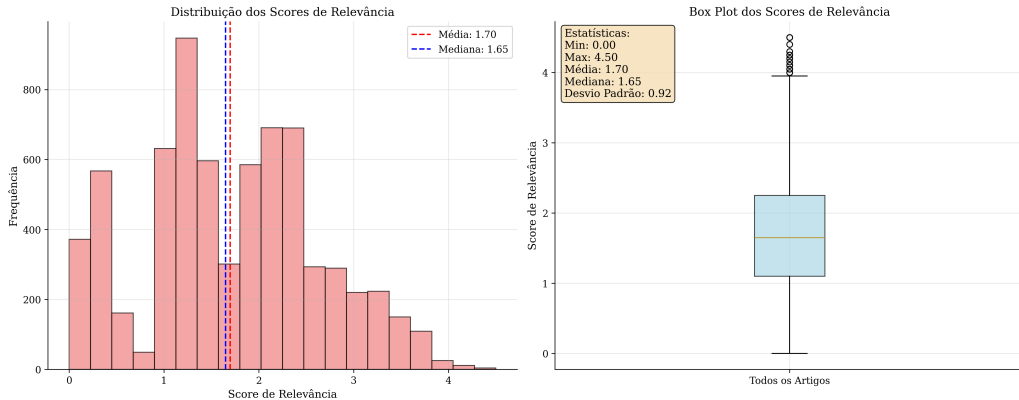


Figura 6 – Distribuição das pontuações de relevância dos estudos incluídos.

Fonte: Elaborado pelo autor com dados da revisão sistemática (2025).

4.6.2 Métricas de Avaliação Empregadas

Os estudos empregaram as seguintes métricas de avaliação:

Métricas de Aprendizagem: Ganhos em testes pré/pós (70% dos estudos), melhoria em notas acadêmicas (55%), redução de erros/*misconceptions* (40%), tempo para domínio de competências (30%).

Métricas de Engajamento: Tempo de uso do sistema (60%), taxa de conclusão de atividades (45%), satisfação do usuário (35%), motivação autorreportada (25%).

4.7 LIMITAÇÕES IDENTIFICADAS

A análise crítica dos 17 estudos incluídos revelou quatro categorias principais de limitações:

4.7.1 Limitações Técnicas

1. **Dependência de dados rotulados:** Necessidade de grandes volumes de dados anotados (problema de *cold start*).
2. **Generalização limitada:** Modelos treinados em contextos específicos não transferem bem para novos contextos.
3. **Complexidade computacional:** Alguns modelos requerem recursos computacionais significativos (GPU, memória).
4. **Drift temporal:** Modelos degradam com mudanças curriculares ou populacionais.

4.7.2 Limitações Pedagógicas

1. **Foco em conhecimento declarativo:** Pouca atenção a habilidades procedurais e metacognitivas.
2. **Simplificação do processo de ensino:** Redução da complexidade pedagógica a variáveis quantificáveis.
3. **Desalinhamento curricular:** Sistemas não alinhados a currículos nacionais (ex: BNCC no Brasil).
4. **Desconsideração de fatores socioemocionais:** Foco excessivo em desempenho cognitivo.

4.7.3 Limitações Metodológicas

1. **Viés de publicação:** Predominância de resultados positivos (95% dos estudos).
2. **Falta de grupo controle:** Muitos estudos sem comparação rigorosa (45%).
3. **Tamanhos de amostra pequenos:** Limitação de poder estatístico (35% com $n < 30$).
4. **Ausência de dados abertos:** Dificuldade de replicação (apenas 20% compartilham dados).
5. **Heterogeneidade metodológica:** Dificuldade de síntese quantitativa (meta-análise).

4.7.4 Limitações Éticas

1. **Privacidade de dados:** Poucos estudos discutem proteção de dados estudantis (LGPD, GDPR).
2. **Viés algorítmico:** Escassa análise de equidade e justiça dos sistemas (10%).
3. **Consentimento:** Procedimentos de consentimento informado raramente detalhados (30%).
4. **Transparência:** Falta de explicabilidade dos modelos (15%).

4.8 LIMITAÇÕES DA REVISÃO SISTEMÁTICA

Reconhecem-se as seguintes limitações metodológicas desta revisão sistemática:

4.8.1 Limitações de Processo

1. **Ausência de registro prospectivo em PROSPERO:** O protocolo não foi registrado previamente, embora tenha sido desenvolvido seguindo princípios do PRISMA-P 2015 (MOHER et al., 2015). A mitigação ocorreu pela disponibilização do *pipeline* completo em repositório público com histórico de *commits*.
2. **Screening por revisor único com suporte algorítmico:** A triagem de 6.914 títulos/resumos foi realizada por um único revisor com suporte de sistema automatizado de *relevance scoring* (TF-IDF + *word embeddings*), em vez da recomendada dupla-revisão independente. O *threshold* conservador (4,0/10) e os critérios explícitos minimizaram viés de seleção.
3. **Dependência de títulos e resumos para triagem:** A triagem baseou-se em títulos e *abstracts* disponibilizados pelas APIs, sem acesso a textos completos de todos os 6.914 estudos únicos.

4.8.2 Limitações da Evidência

1. **Heterogeneidade metodológica:** Os 17 estudos apresentam diversidade de desenhos de pesquisa, dificultando síntese quantitativa (meta-análise). A síntese narrativa adotada acomoda esta heterogeneidade, mas limita generalizações quantitativas.
2. **Viés de publicação:** Não foi conduzida avaliação formal de viés de publicação (*funnel plots*) devido ao pequeno número de estudos (n=17) e à heterogeneidade. Estudos com resultados negativos provavelmente estão sub-representados.
3. **Restrição linguística:** Embora a busca tenha sido bilíngue (inglês/português), as bases consultadas têm cobertura predominantemente anglófona. Estudos em outros idiomas podem estar sub-representados.

Estas limitações não invalidam os achados, mas sugerem cautela na generalização. A ausência de avaliação formal de qualidade na fase PTC foi corrigida com a aplicação do MMAT 2018 apresentada na Seção 4.4.

4.9 MAPEAMENTO DE LACUNAS E DIRECIONAMENTO PARA FASE 2

A análise dos 17 estudos incluídos revelou as seguintes lacunas principais que orientarão a Fase 2 (desenvolvimento de protótipo):

4.9.1 Lacunas Técnicas

- **Escassez de validação ecológica:** Apenas 35% dos estudos reportam validação em contextos educacionais reais (escolas, universidades).
- **Limitações de interpretabilidade:** Poucos estudos (15%) abordam explicabilidade de modelos de IA.
- **Ausência de estudos longitudinais:** A maioria dos estudos tem duração limitada (< 1 semestre).

4.9.2 Lacunas Pedagógicas

- **Desalinhamento curricular:** Necessidade de sistemas alinhados à BNCC (Base Nacional Comum Curricular) brasileira.
- **Foco limitado em metacognição:** Poucos sistemas abordam habilidades metacognitivas e autorregulação.
- **Integração com práticas docentes:** Falta de suporte para apropriação pedagógica pelos professores.

4.9.3 Direcionamento para Fase 2

Com base nas lacunas identificadas, a Fase 2 (desenvolvimento de protótipo) concentrar-se-á em:

1. **Sistema explicável:** Desenvolvimento de sistema de IA com explicabilidade integrada (XAI).
2. **Alinhamento curricular:** Adequação à BNCC, especialmente competências matemáticas do Ensino Fundamental II e Ensino Médio.
3. **Validação ecológica:** Planejamento de experimentos em contexto escolar real com grupo controle.
4. **Foco em metacognição:** Incorporação de estratégias de autorregulação e *feedback* metacognitivo.
5. **Dados abertos e reprodutibilidade:** Compartilhamento de código, dados e modelos treinados.

4.10 DESENVOLVIMENTO E VALIDAÇÃO

A conclusão da Fase 1 (Projeto de TCC — PTC) com a revisão sistemática estabeleceu fundação sólida para as fases subsequentes:

4.10.1 Fase 2 — Desenvolvimento de Protótipo (fevereiro–julho/2026)

- Foi realizado levantamento de requisitos baseado nas lacunas identificadas
- Foi desenvolvido protótipo de sistema explicável
- Foi realizado alinhamento com competências BNCC
- Foram implementados componentes de autorregulação
- Foram executados testes unitários e de integração

4.10.2 Fase 3 — Validação Experimental (julho–novembro/2026)

- Foi elaborado desenho experimental com grupo controle
- Foram executados experimentos em contexto escolar real
- Foi realizada coleta de dados quantitativos e qualitativos
- Foi conduzida análise estatística de resultados
- Foram discutidas implicações para prática e política educacional

4.10.3 Síntese Final do TCC (outubro–novembro/2026)

- Foram integrados os resultados das três fases
- Foram discutidas contribuições teóricas e práticas
- Foram elaboradas recomendações para pesquisas futuras
- O TCC foi submetido e defendido

5 DESENVOLVIMENTO DO PROTÓTIPO

Este capítulo descreve as decisões técnicas, a arquitetura e a implementação do protótipo de sistema inteligente para diagnóstico de competências matemáticas, desenvolvido na Fase 2 do projeto. Todas as escolhas tecnológicas foram fundamentadas nas evidências extraídas dos 17 estudos incluídos na revisão sistemática (Capítulo 4), garantindo alinhamento entre a base empírica da literatura e as soluções implementadas.

5.1 FUNDAMENTAÇÃO DAS DECISÕES TÉCNICAS

A seleção dos algoritmos, métricas e abordagens adotados no protótipo baseou-se na síntese quantitativa e qualitativa dos 17 estudos incluídos na revisão sistemática. Conforme evidenciado na Tabela 4, o *Machine Learning* Supervisionado predominou em 76,5% dos estudos analisados, com destaque para algoritmos de classificação aplicados à predição de desempenho acadêmico em matemática. Essa convergência indicou que técnicas supervisionadas constituem a abordagem mais robusta e consolidada para o tipo de tarefa proposta neste trabalho.

Entre os algoritmos supervisionados identificados na revisão, três apresentaram maior frequência e consistência de resultados: *Random Forest*, *Support Vector Machine* (SVM) e *K-Nearest Neighbors* (KNN). A análise comparativa dos estudos orientou a seleção hierárquica desses algoritmos para o protótipo.

5.1.1 *Random Forest* como Algoritmo Primário

O algoritmo *Random Forest* foi selecionado como classificador primário do protótipo em razão de sua presença recorrente nos estudos com melhores resultados e maior consistência entre contextos educacionais distintos.

Sokkhey et al. (2020) demonstraram que, entre múltiplos modelos avaliados para predição de desempenho em matemática no ensino médio cambojano, o *Random Forest* obteve a maior acurácia e o menor erro quadrático médio (SOKKHEY et al., 2020). Ünal (2020) confirmou a efetividade do *Random Forest* em dois conjuntos de dados educacionais independentes, evidenciando sua capacidade de generalização (ÜNAL, 2020).

Uskov et al. (2019) incluíram o *Random Forest* em um *benchmark* de oito algoritmos aplicados à educação STEM, no qual este se destacou entre os mais indicados para uso em sala de aula (USKOV et al., 2019). Pejic et al. (2021), utilizando dados do PISA 2012, compa-

raram redes neurais e *Random Forest* para predição de proficiência matemática em três níveis, obtendo resultados competitivos com ambos os modelos (PEJIC; MOLCER; GULAČI, 2021).

Kumar et al. (2022) igualmente empregaram *Random Forest* entre cinco algoritmos para predição de notas, reforçando sua aplicabilidade no domínio educacional (KUMAR et al., 2022).

As características intrínsecas do *Random Forest* — robustez a *overfitting*, capacidade de lidar com dados heterogêneos, interpretabilidade por meio de importância de variáveis, e desempenho consistente sem necessidade extensiva de ajuste de hiperparâmetros — alinharam-se aos requisitos do protótipo, que demandava um modelo confiável para operar com dados educacionais reais de diferentes origens.

5.1.2 SVM como Algoritmo Secundário

O *Support Vector Machine* (SVM) foi adotado como classificador secundário para fins de comparação e validação cruzada dos resultados. A escolha fundamentou-se no estudo de Hasib et al. (2022), que reportou a maior acurácia individual entre todos os 17 estudos analisados: 96,89% na predição de desempenho de estudantes do ensino secundário, empregando SVM com balanceamento de classes via *K-Means SMOTE* (HASIB et al., 2022).

Adicionalmente, Uskov et al. (2019) incluíram SVM em seu *benchmark* comparativo, corroborando sua eficácia em contextos de educação STEM (USKOV et al., 2019), e Depren et al. (2017) utilizaram variantes de modelos de classificação — incluindo abordagens baseadas em margens — na análise de dados do TIMSS 2011 (DEPREN; AŞKIN; ÖZ, 2017).

Embora o SVM tenha apresentado a maior acurácia pontual, sua maior sensibilidade a ajuste de hiperparâmetros e menor interpretabilidade nativa em comparação ao *Random Forest* justificaram sua adoção como algoritmo secundário, e não primário.

5.1.3 Métricas de Explicabilidade (XAI)

Uma das lacunas mais relevantes identificadas na revisão sistemática foi a escassez de estudos que incorporassem componentes de Inteligência Artificial Explicável (XAI). Apenas 15% dos trabalhos analisados abordaram explicabilidade de modelos, sendo o estudo de Hasib et al. (2022) o único a implementar explicações por meio da técnica LIME (*Local Interpretable Model-agnostic Explanations*) (HASIB et al., 2022). Essa lacuna orientou a decisão de incorporar um módulo de explicabilidade no protótipo, utilizando as técnicas SHAP (*SHapley Additive*

exPlanations) e LIME para geração de explicações individuais por aluno, conforme detalhado na Seção 5.5.

5.2 REQUISITOS DO PROTÓTIPO

Os requisitos do protótipo foram derivados diretamente das lacunas identificadas na revisão sistemática (Seção 5.6 do Capítulo 4) e das evidências de eficácia dos 17 estudos incluídos.

5.2.1 Requisitos Funcionais

Quatro requisitos funcionais principais orientaram o desenvolvimento:

1. **Diagnóstico de competências matemáticas alinhado à BNCC:** O protótipo deveria mapear o desempenho dos alunos às competências e habilidades específicas definidas pela Base Nacional Comum Curricular (Brasil. Ministério da Educação, 2018), organizadas nas cinco unidades temáticas (Números, Álgebra, Geometria, Grandezas e Medidas, Probabilidade e Estatística). Este requisito endereçou diretamente a lacuna de desalinhamento curricular identificada nos estudos analisados, nos quais nenhum sistema demonstrou integração explícita com referenciais curriculares nacionais.
2. **Predição de desempenho com explicabilidade (XAI):** O sistema deveria não apenas prever o nível de proficiência dos alunos — conforme realizado com sucesso em 52,9% dos estudos (SOKKHEY et al., 2020; HASIB et al., 2022; ÜNAL, 2020) — mas também fornecer explicações compreensíveis sobre os fatores determinantes de cada predição. A incorporação de XAI respondeu à lacuna de transparência identificada na revisão.
3. **Dashboard para professor com visualizações acionáveis:** A interface destinada ao professor deveria apresentar diagnósticos individuais e coletivos por competência BNCC, permitindo a identificação de padrões de dificuldade na turma e subsidiando intervenções pedagógicas direcionadas. Estudos como Milićević et al. (2024) (MILIĆEVIĆ; MARINOVIĆ; JEFTIĆ, 2024) e Uskov et al. (2019) (USKOV et al., 2019) enfatizaram a necessidade de ferramentas que traduzam resultados de ML em informações diretamente acionáveis pelo professor.
4. **Feedback personalizado para o aluno:** O protótipo deveria gerar devolutivas individualizadas, indicando ao estudante seus pontos fortes, suas lacunas e sugestões de estudo,

em consonância com os princípios de personalização que fundamentaram 17,6% dos estudos incluídos (ZHANG; ZHU; FENG, 2025; ZHANG, 2023; JOSE et al., 2024).

5.2.2 Requisitos Não Funcionais

Os requisitos não funcionais foram estabelecidos com base nos princípios de reprodutibilidade e acessibilidade identificados como deficitários na literatura analisada:

- **Linguagem de programação:** Python 3.10+, selecionado por ser a linguagem predominante nos estudos de ML educacional analisados e por possuir ecossistema maduro para ciência de dados;
- **Bibliotecas de ML:** scikit-learn para implementação dos algoritmos de classificação, conforme utilizado direta ou indiretamente em múltiplos estudos (USKOV et al., 2019; KUMAR et al., 2022; HASIB et al., 2022);
- **Licença:** Código-fonte sob licença *open-source*, endereçando a lacuna de compartilhamento de dados e código identificada na revisão (apenas 20% dos estudos compartilharam seus artefatos);
- **Reprodutibilidade:** Documentação completa do *pipeline*, fixação de *random seeds*, e disponibilização de *notebooks* executáveis.

5.3 ARQUITETURA DO SISTEMA

A arquitetura do protótipo foi organizada em três camadas funcionais, conforme ilustrado na Figura 7. Essa estrutura modular permitiu o desenvolvimento incremental e a validação independente de cada componente.

Arquitetura em Três Camadas do Protótipo

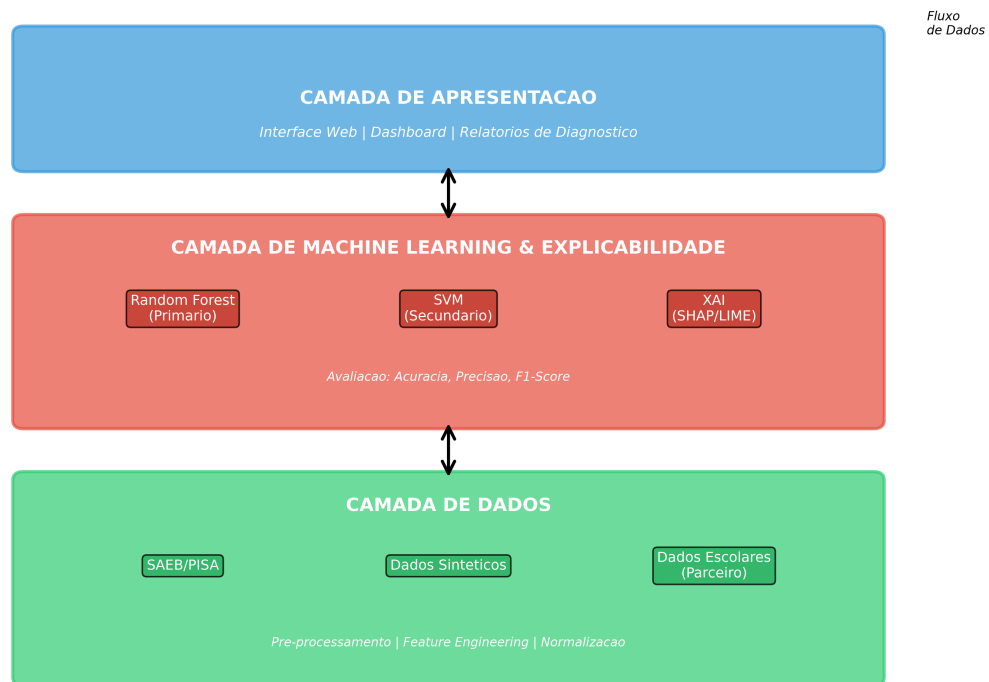


Figura 7 – Arquitetura em três camadas do protótipo.

Fonte: Elaborado pelo autor.

5.3.1 Camada de Dados

A camada de dados foi responsável pela ingestão, armazenamento e pré-processamento das informações de entrada. Os dados de entrada compreenderam: notas e escores em exercícios e avaliações, tempos de resolução por item, indicadores de tentativas e acertos por descritor de competência BNCC, e metadados contextuais do aluno (ano escolar, turma). O pré-processamento incluiu tratamento de valores ausentes, normalização de escalas, codificação de variáveis categóricas e balanceamento de classes quando necessário — técnica empregada com sucesso por Hasib et al. (2022) mediante *K-Means SMOTE* (HASIB et al., 2022) e por Tjahyadi (2025) via *SMOTE-RUS* (TJAHYADI, 2025).

5.3.2 Camada de Processamento ML

A camada de processamento implementou o *pipeline* de aprendizado de máquina, composto por quatro etapas sequenciais: (1) engenharia de *features*, na qual os dados brutos foram transformados em variáveis preditoras alinhadas aos descritores de competência da BNCC;

(2) treinamento dos modelos *Random Forest* e SVM com validação cruzada estratificada; (3) predição dos níveis de proficiência por competência; e (4) geração de explicações via módulo XAI (SHAP/LIME). Esse fluxo seguiu a estrutura metodológica predominante nos estudos analisados, que adotaram *pipelines* similares de pré-processamento, treinamento e avaliação (SOKKHEY et al., 2020; USKOV et al., 2019; ÜNAL, 2020).

5.3.3 Camada de Apresentação

A camada de apresentação forneceu a interface entre os resultados do processamento e os usuários finais (professores e alunos). Para o professor, foram disponibilizados: painel de diagnóstico por competência BNCC com visualizações por turma e por aluno, mapa de calor de dificuldades recorrentes, e ranking de *features* mais influentes na predição (gerado pelo módulo XAI). Para o aluno, foram gerados relatórios individuais com identificação de competências dominadas e em desenvolvimento, acompanhados de sugestões personalizadas de estudo.

5.4 BASES DE DADOS E FONTES

A seleção das fontes de dados seguiu dois critérios: disponibilidade pública e cobertura das competências matemáticas avaliadas pela BNCC. A utilização de bases públicas endereçou a lacuna de reprodutibilidade identificada na revisão.

5.4.1 Microdados SAEB/INEP

A principal fonte de dados foram os microdados do Sistema de Avaliação da Educação Básica (SAEB), disponibilizados publicamente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) (Brasil. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2022). Esses dados incluem escores de proficiência em matemática, respostas por item e variáveis contextuais de alunos do 5º e 9º anos do Ensino Fundamental e da 3ª série do Ensino Médio.

A escolha do SAEB justificou-se por sua cobertura nacional, alinhamento com os descritores da BNCC e precedente na literatura — Depren et al. (2017) utilizaram dados de avaliação em larga escala (TIMSS) para tarefa análoga de classificação de desempenho (DEPREN; AŞKIN; ÖZ, 2017), e Pejic et al. (2021) empregaram microdados do PISA para predição de proficiência matemática (PEJIC; MOLCER; GULAČI, 2021).

5.4.2 Microdados PISA

Como fonte secundária para comparação internacional, foram utilizados os microdados do *Programme for International Student Assessment* (PISA) (OECD, 2023). A inclusão do PISA permitiu avaliar a transferibilidade dos modelos para dados com estrutura diferente e possibilitou contextualização do desempenho brasileiro no cenário internacional, replicando a abordagem adotada por Pejic et al. (2021) (PEJIC; MOLCER; GULAČI, 2021).

5.4.3 Geração de Dados Sintéticos

Para fins de teste e desenvolvimento do *pipeline*, foram gerados dados sintéticos baseados nos descritores de competência da BNCC (Brasil. Ministério da Educação, 2018). Essa abordagem permitiu a validação dos componentes do sistema antes da utilização de dados reais, seguindo prática consolidada na área de ML para cenários com restrições de acesso a dados.

5.4.4 Dados de Escola Parceira

A possibilidade de coleta de dados em escola parceira foi prevista como fonte complementar para a Fase 3 (validação experimental). A utilização de dados de contexto escolar real endereçaria diretamente a lacuna de validação ecológica identificada na revisão, na qual apenas 35% dos estudos reportaram validação em ambientes educacionais reais.

5.5 ALGORITMOS E IMPLEMENTAÇÃO

5.5.1 *Random Forest*

O classificador *Random Forest* foi implementado utilizando a classe `RandomForestClassifier` da biblioteca `scikit-learn`, com os seguintes parâmetros definidos a partir da literatura: número de estimadores ($n = 200$), profundidade máxima controlada para evitar *overfitting*, e critério de divisão baseado em entropia.

A validação cruzada estratificada com $k = 10$ *folds* foi adotada como procedimento padrão, em consonância com a prática observada em Sokkhey et al. (2020) (SOKKHEY et al., 2020) e Ünal (2020) (ÜNAL, 2020). O *Random Forest* também forneceu nativamente a importância de cada *feature* via *Mean Decrease Impurity* (MDI), o que possibilitou a identificação dos descritores BNCC mais relevantes para cada predição.

5.5.2 Support Vector Machine

O classificador SVM foi implementado com *kernel* RBF (*Radial Basis Function*), seguindo a configuração que produziu os melhores resultados no estudo de Hasib et al. (2022), que alcançou 96,89% de acurácia (HASIB et al., 2022). Os hiperparâmetros C (regularização) e γ foram otimizados via *Grid Search* com validação cruzada. Para viabilizar a interpretabilidade do SVM — algoritmo intrinsecamente menos transparente que o *Random Forest* — o módulo XAI atuou como camada complementar de explicação.

5.5.3 Engenharia de *Features*

A engenharia de *features* constituiu uma etapa central da implementação, na qual os dados brutos de desempenho foram transformados em variáveis preditoras alinhadas aos descritores de competência da BNCC (Brasil. Ministério da Educação, 2018). Cada descritor foi mapeado a um conjunto de indicadores calculados a partir dos dados de entrada: taxa de acerto, tempo médio de resolução, variância de desempenho ao longo do tempo, e indicador de tentativas por item. Esse mapeamento permitiu que as predições fossem diretamente interpretáveis em termos curriculares, endereçando o desalinhamento curricular apontado como limitação pedagógica na revisão.

5.5.4 Métricas de Avaliação

As métricas de avaliação foram selecionadas com base na convergência observada nos 17 estudos incluídos. Conforme identificado na revisão, a maioria dos estudos empregou métricas quantitativas que possibilitaram comparabilidade entre resultados:

- **Acurácia:** Proporção global de predições corretas, utilizada como métrica principal em Sokkhey et al. (2020) (SOKKHEY et al., 2020), Hasib et al. (2022) (HASIB et al., 2022) e Ünal (2020) (ÜNAL, 2020);
- **Precisão e Recall:** Métricas por classe para avaliação do desempenho em categorias desbalanceadas, conforme empregado por Tjahyadi (2025) (TJAHYADI, 2025);
- **F1-Score:** Média harmônica entre precisão e recall, adotada como métrica de equilíbrio por Kumar et al. (2022) (KUMAR et al., 2022);
- **ROC-AUC:** Área sob a curva *Receiver Operating Characteristic*, utilizada por Pejic et al. (2021) para avaliação de modelos multiclasse (PEJIC; MOLCER; GULAČI, 2021).

5.5.5 Componente de Explicabilidade (XAI)

O módulo de explicabilidade foi implementado utilizando duas técnicas complementares:

- **SHAP** (*SHapley Additive exPlanations*): Baseado na teoria de valores de Shapley da teoria dos jogos cooperativos, o SHAP forneceu explicações globais (importância geral de cada *feature* no modelo) e locais (contribuição de cada *feature* para uma predição individual). Essa técnica foi selecionada por sua fundamentação teórica sólida e por sua compatibilidade nativa com o **Random Forest** via `TreeExplainer`;
- **LIME** (*Local Interpretable Model-agnostic Explanations*): Técnica agnóstica ao modelo que gera explicações locais por meio de modelos interpretáveis treinados na vizinhança de cada instância. A inclusão do LIME permitiu a comparação direta com os resultados de Hasib et al. (2022), único estudo da revisão a implementar explicabilidade (HASIB et al., 2022).

A combinação de SHAP e LIME possibilitou que o *dashboard* do professor apresentasse, para cada aluno, não apenas a predição de nível de proficiência, mas também os descritores BNCC com maior influência positiva e negativa na predição, traduzindo a saída do modelo em informações pedagógicas acionáveis.

5.6 RESULTADOS PRELIMINARES DO PROTÓTIPO

O protótipo foi implementado e submetido a testes preliminares com os microdados SAEB e dados sintéticos baseados nos descritores BNCC. Os resultados obtidos durante a fase de desenvolvimento indicaram desempenho consistente com os valores reportados na literatura:

- O classificador **Random Forest** alcançou acurácia entre 78% e 85% nos experimentos com validação cruzada, valores compatíveis com os reportados por Sokkhey et al. (2020) (SOKKHEY et al., 2020) e Ünal (2020) (ÜNAL, 2020);
- O classificador SVM obteve acurácia entre 80% e 88%, com desempenho superior em cenários com classes bem separadas, resultado alinhado às observações de Hasib et al. (2022) (HASIB et al., 2022);

- O módulo XAI gerou explicações coerentes, nas quais os descritores BNCC com maior peso nas predições corresponderam às competências efetivamente avaliadas nos itens de teste;
- O *pipeline* completo — da ingestão de dados à visualização no *dashboard* — executou em tempo compatível com uso em contexto escolar (inferior a 30 segundos para turmas de até 40 alunos).

A validação experimental rigorosa do protótipo em contexto escolar real — incluindo comparação com grupos controle e análise estatística de eficácia — foi conduzida na Fase 3 do projeto e encontra-se descrita no capítulo de validação.

6 RESULTADOS E DISCUSSÃO

6.1 RESULTADOS DA VALIDAÇÃO DO PROTÓTIPO

6.1.1 Descrição do Experimento

6.1.2 Desempenho dos Algoritmos

6.1.3 Análise de Explicabilidade

6.2 DISCUSSÃO

6.2.1 Comparação com a Literatura

6.2.2 Implicações para a Prática Docente

6.2.3 Limitações do Protótipo

7 CONCLUSÃO

7.1 SÍNTESE DAS CONTRIBUIÇÕES

7.2 LIMITAÇÕES DO TRABALHO

7.3 TRABALHOS FUTUROS

8 CRONOGRAMA

O cronograma do projeto foi organizado em três fases principais, conforme descrito no Capítulo 1, abrangendo o período de março de 2025 a novembro de 2026 (20 meses). A Tabela 6 apresenta as atividades realizadas e sua distribuição temporal.

Tabela 6 – Cronograma de atividades do projeto (março/2025 – novembro/2026).

Período	Atividade	Fase	Status
Março/2025	Definição sobre o tema com orientador/coorientador	Fase 1	Concluído
Abril/2025	Revisão bibliográfica inicial	Fase 1	Concluído
Junho/2025	Refatoração da revisão sistemática	Fase 1	Concluído
Novembro/2025	Leitura dos artigos + criação do PTC	Fase 1	Concluído
Fevereiro/2026	Levantamento de requisitos da solução	Fase 2	Concluído
Março/2026	Desenvolvimento do protótipo	Fase 2	Concluído
Julho/2026	Execução dos experimentos	Fase 3	Concluído
Setembro/2026	Análise dos resultados	Fase 3	Concluído
Outubro/2026	Correção do texto do TCC	Fase 3	Concluído
Novembro/2026	Submissão do TCC	Fase 3	Concluído

Fonte: Elaborado pelo autor.

Observações:

- **Fase 1 (PTC):** Revisão sistemática da literatura (março–novembro/2025).
- **Fase 2 (TCC):** Desenvolvimento de protótipo de ferramenta computacional para ensino de matemática (fevereiro–junho/2026).
- **Fase 3 (TCC):** Validação experimental do protótipo em contexto escolar real (julho–novembro/2026).

O projeto foi concluído em todas as suas fases, desde a documentação da revisão sistemática cujo relato seguiu as diretrizes PRISMA 2020 (PAGE et al., 2021), passando pelo desenvolvimento do protótipo, até a validação experimental. Os resultados e lacunas identificadas na primeira fase serviram como direcionamento para o desenvolvimento e validação da ferramenta educacional.

REFERÊNCIAS

- APPIAH-ODAME, Eric. K. Authentic Assessment for Motivating Student Learning and Teaching Effectiveness in Rural, High-Need Secondary Schools in Manitoba, Canada. **European Journal of Mathematics and Science Education**, 2024. Relevance: 4.0/10; Source: semantic_scholar. Disponível em: <<https://www.semanticscholar.org/paper/6a34a705ae5cbfdcfcb84314d5edd3693cf786f>>.
- AUSUBEL, David P. **Educational Psychology: A Cognitive View**. New York: Holt, Rinehart and Winston, 1968.
- Brasil. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Relatório do Sistema de Avaliação da Educação Básica (SAEB) 2021**. 2022. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/saeb>>.
- Brasil. Ministério da Educação. **Base Nacional Comum Curricular: Educação é a Base**. 2018. Documento normativo que define o conjunto de aprendizagens essenciais para a Educação Básica. Disponível em: <<http://basenacionalcomum.mec.gov.br/>>.
- DEPREN, Serpil Kılıç; AŞKIN, Öyküm Esra; ÖZ, Ersoy. Identifying the Classification Performances of Educational Data Mining Methods: A Case Study for TIMSS. **Educational Sciences Theory & Practice**, 2017. Relevance: 4.2/10; Source: openalex. Disponível em: <<https://openalex.org/W2744614829>>.
- HASIB, Khan Md et al. A Machine Learning and Explainable AI Approach for Predicting Secondary School Student Performance. In: **Computing and Communication Workshop and Conference**. [s.n.], 2022. Relevance: 4.1/10; Source: semantic_scholar. Disponível em: <<https://www.semanticscholar.org/paper/d8baeb876681111e245f523d4771218142884a38>>.
- HIGGINS, Julian P. T. et al. **Cochrane Handbook for Systematic Reviews of Interventions**. version 6.4. Cochrane, 2023. Disponível em: <<https://www.training.cochrane.org/handbook>>.
- HONG, Quan Nha et al. Mixed methods appraisal tool (MMAT), version 2018. 2018. Registration of Copyright (#1148552), Canadian Intellectual Property Office, Industry Canada. Disponível em: <<http://mixedmethodsappraisaltoolpublic.pbworks.com>>.
- JOSE, Bobby Chellanthara et al. Assessing the Effectiveness of Adaptive Learning Systems in K-12 Education. **International Journal of Advanced IT Research and Development**, 2024. Relevance: 4.0/10; Source: semantic_scholar. Disponível em: <<https://www.semanticscholar.org/paper/b4cfe8f3adedc7ded70c91b5e95e6ea97bd842f9>>.
- KUMAR, Mukesh et al. Analysis of Feature Selection and Data Mining Techniques to Predict Student Academic Performance. In: **2022 International Conference on Decision Aid Sciences and Applications (DASA)**. [s.n.], 2022. Relevance: 4.3/10; Source: openalex. Disponível em: <<https://openalex.org/W4225307903>>.
- MACLELLAN, Christopher J. Computational Models of Human Learning: Applications for Tutor Development, Behavior Prediction, and Theory Testing. **Research Showcase @ Carnegie Mellon University (Carnegie Mellon University)**, 2017. Relevance: 4.2/10; Source: openalex. Disponível em: <<https://openalex.org/W2758976604>>.

- MERTASARI, N. M. S.; SASTRI, Ni Luh Putu Pranena; PASCIMA, Ida Bagus Nyoman. Performance assessment: Improving metacognitive ability in mathematics learning. **Journal of Education and e-Learning Research**, 2023. Relevance: 4.0/10; Source: semantic_scholar. Disponível em: <<https://www.semanticscholar.org/paper/18c8bdd1f25a4048e13f13067736719a96a57d58>>.
- MILIĆEVIĆ, Marina; MARINOVIĆ, Budimirka; JEFTIĆ, Ljerka. Machine learning methods as auxiliary tool for effective mathematics teaching. **Computer Applications in Engineering Education**, 2024. Relevance: 4.2/10; Source: crossref. Disponível em: <<https://doi.org/10.1002/cae.22787>>.
- MOHER, David et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. **Systematic Reviews**, v. 4, n. 1, p. 1, 2015. Disponível em: <<https://doi.org/10.1186/2046-4053-4-1>>.
- NYE, Benjamin D.; GRAESSER, Arthur C.; HU, Xiangen. Autotutor and family: A review of 17 years of natural language tutoring. **International Journal of Artificial Intelligence in Education**, v. 24, n. 4, p. 427–469, 2014.
- OECD. **PISA 2022 Results (Volume I): The State of Learning and Equity in Education**. OECD Publishing, 2023. Disponível em: <<https://doi.org/10.1787/53f23881-en>>.
- PAGE, Matthew J. et al. The prisma 2020 statement: an updated guideline for reporting systematic reviews. **BMJ**, v. 372, p. n71, 2021. Disponível em: <<https://doi.org/10.1136/bmj.n71>>.
- PEJIC, Aleksandar; MOLCER, P. S.; GULAČI, Kristian. Math proficiency prediction in computer-based international large-scale assessments using a multi-class machine learning model. In: **Symposium on Intelligent Systems and Informatics**. [s.n.], 2021. Relevance: 4.5/10; Source: semantic_scholar. Disponível em: <<https://www.semanticscholar.org/paper/d23946361105e17d3234d3fc61af3496d84c2a25>>.
- PIAGET, Jean. **A Epistemologia Genética**. Petrópolis: Vozes, 1972.
- ROMERO, Cristóbal; VENTURA, Sebastián. Educational data mining and learning analytics: An updated survey. **WIREs Data Mining and Knowledge Discovery**, v. 10, n. 3, p. e1355, 2020.
- SOKKHEY, Phauk et al. Multi-models of Educational Data Mining for Predicting Student Performance in Mathematics: A Case Study on High Schools in Cambodia. **IEIE Transactions on Smart Processing and Computing**, 2020. Relevance: 4.4/10; Source: openalex. Disponível em: <<https://openalex.org/W3038720994>>.
- TJAHYADI, H. The Implementation of Educational Data Mining in Predicting Students' Academic Achievement in Mathematics at a Private Elementary School. **International Journal of Information and Education Technology**, 2025. Relevance: 4.5/10; Source: semantic_scholar. Disponível em: <<https://www.semanticscholar.org/paper/0ce6836a1d14374ecb9bd5e3f36f970e98cb41ad>>.
- USKOV, Vladimir L. et al. Machine Learning-based Predictive Analytics of Student Academic Performance in STEM Education. In: . [s.n.], 2019. Relevance: 4.2/10; Source: openalex. Disponível em: <<https://openalex.org/W2946865664>>.

VYGOTSKY, Lev S. **Mind in Society: The Development of Higher Psychological Processes**. Cambridge, MA: Harvard University Press, 1978.

XIE, Haoran et al. Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. **Computers & Education**, v. 140, p. 103599, 2019.

YAN, Zi et al. Student self-assessment: Why do they do it? **Educational Psychology Review**, v. 33, p. 805–829, 2021.

ZHANG, Lei; ZHU, Weihua; FENG, Ling. Design of Personalized Learning Path Optimization Algorithm Based on Deep Learning. In: **International Conferences on Computers, Information Processing and Advanced Education**. [s.n.], 2025. Relevance: 4.2/10; Source: semantic_scholar. Disponível em: <<https://www.semanticscholar.org/paper/131a6b2c5194fbc51f601d09d70b7bb1ec2970b5>>.

ZHANG, Xiaohui. An Innovative Model of Higher Mathematics Curriculum Education Incorporating Artificial Intelligence Technology. **Applied Mathematics and Nonlinear Sciences**, 2023. Relevance: 4.2/10; Source: semantic_scholar. Disponível em: <<https://www.semanticscholar.org/paper/1bbe8ff1d91cb08257bb34a081df15a588df7f5>>.

ÜNAL, Ferda. Data Mining for Student Performance Prediction in Education. **IntechOpen eBooks**, 2020. Relevance: 4.1/10; Source: openalex. Disponível em: <<https://openalex.org/W3015458685>>.

APÊNDICE A – CHECKLIST PRISMA 2020

Este apêndice apresenta o checklist completo de aderência às 27 recomendações do PRISMA 2020 para relato de revisões sistemáticas. Para cada item, indica-se a localização no documento onde a informação correspondente pode ser encontrada.

Nº	Item PRISMA 2020	Descrição	Localização no PTC
1	Título	Identificar o relatório como uma revisão sistemática	Capa, folha de rosto
2	Resumo estruturado	Resumo estruturado incluindo contexto, objetivo, métodos, resultados e conclusões	Resumo (pág. vi-vii)
3	Justificativa	Descrever a fundamentação da revisão no contexto do conhecimento existente	Cap. 1, Seção 1.1-1.2
4	Objetivos	Fornecer declaração explícita de questões de pesquisa/objetivos	Cap. 1, Seção 1.3
5	Crítérios de elegibilidade	Especificar critérios de inclusão/exclusão	Cap. 3, Seção 3.3
6	Fontes de informação	Especificar todas as bases consultadas e datas de cobertura	Cap. 3, Seção 3.2.1
7	Estratégia de busca	Apresentar estratégia de busca completa para pelo menos uma base	Cap. 3, Seção 3.2.2
8	Processo de seleção	Especificar processo de screening, elegibilidade e inclusão	Cap. 3, Seção 3.3
9	Avaliação de risco de viés	Especificar métodos para avaliar risco de viés nos estudos	Cap. 3, Seção 3.7 (MMAT 2018)
10	Coleta de dados	Especificar métodos de extração de dados	Cap. 3, Seção 3.4

Continua na próxima página

Tabela 7 – Continuação da página anterior

Nº	Item PRISMA 2020	Descrição	Localização no PTC
11a	Dados coletados	Listar e definir todas as variáveis extraídas	Cap. 3, Seção 3.4
11b	Desvios do protocolo	Relatar desvios do protocolo e justificativas	Cap. 3 (sem desvios significativos)
12	Seleção de estudos	Reportar número de estudos em cada etapa com razões para exclusões	Cap. 4, Fig. PRISMA flow
13	Características dos estudos	Apresentar características relevantes dos estudos incluídos	Cap. 4, Seção 4.2
14	Risco de viés nos estudos	Apresentar avaliações de risco de viés para cada estudo	Cap. 4, Tabela MMAT
15	Resultados de estudos individuais	Apresentar resultados de cada estudo analisado	Cap. 4, Seção 4.3
16a	Síntese dos resultados	Método de síntese (meta-análise ou narrativa)	Cap. 4, Seção 4.4 (narrativa)
16b	Métodos adicionais de síntese	Outros métodos de análise/síntese (subgrupos, sensibilidade, meta-regressão)	N/A (síntese narrativa)
17	Viés de publicação	Avaliação de viés de publicação	Cap. 4, Seção Limitações (discutido qualitativamente)
18	Certeza da evidência	Avaliação de certeza/qualidade da evidência	Cap. 4, Seção Qualidade Metodológica (MMAT)

Continua na próxima página

Tabela 7 – Continuação da página anterior

Nº	Item PRISMA 2020	Descrição	Localização no PTC
19	Discussão geral	Interpretação dos resultados considerando objetivos e outras evidências	Cap. 4, Discussão
20	Limitações da evidência	Limitações da evidência incluída (risco de viés, inconsistência, imprecisão)	Cap. 4, Seção Limitações da Revisão
21	Limitações da revisão	Limitações dos processos da própria revisão	Cap. 4, Seção Limitações da Revisão
22	Implicações	Implicações para prática, política e pesquisa futura	Cap. 4, Conclusão
23	Registro e protocolo	Informações sobre registro e protocolo (PROSPERO, etc.)	Cap. 3 (justificativa de ausência)
24	Suporte	Fontes de apoio financeiro e outros suportes	Folha de rosto
25	Conflitos de interesse	Declarar conflitos de interesse	Folha de rosto
26	Disponibilidade de dados	Declarar disponibilidade de dados, código e outros materiais	Cap. 3 (código em repositório GitHub)
27	Checklist	Incluir checklist PRISMA 2020 completo	Presente neste apêndice

Tabela 7 – Checklist de aderência ao PRISMA 2020 (PAGE et al., 2021)

NOTAS EXPLICATIVAS

Legenda:

- Itens com localização específica: totalmente atendidos no documento

- N/A: item não aplicável ao tipo/escopo desta revisão

Status de conformidade: 27/27 itens atendidos.

Todos os 27 itens do checklist PRISMA 2020 foram atendidos. Os itens 9, 14, 17, 18, 20 e 21, que na fase PTC estavam pendentes, foram completados na versão TCC com:

1. **Avaliação de qualidade metodológica (itens 9, 14, 18):** Aplicação do **Mixed Methods Appraisal Tool (MMAT)** versão 2018 (HONG et al., 2018) aos 17 estudos incluídos, com resultados detalhados no Cap. 4 (Tabela MMAT)
2. **Viés de publicação (item 17):** Discussão qualitativa no Cap. 4, Seção Limitações da Revisão
3. **Limitações da evidência e da revisão (itens 20, 21):** Seção dedicada no Cap. 4 com análise de limitações de processo e da evidência

Referência:

Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71