

MODELAGEM PREDITIVA BASEADA EM MACHINE LEARNING PARA PREVENÇÃO DA EVASÃO ESCOLAR: UMA REVISÃO SISTEMÁTICA

Priscila Corrêa Cavalcanti
Centro Universitário AlfaUnipac
priscilac.c.fisio@gmail.com

Rafael Pereira Luiz
Centro Universitário AlfaUnipac
rafaelluiz286@gmail.com

Rejane Goecking B.Pereira
Centro Universitário AlfaUnipac
rejanegoecking@hotmail.com

RITA DE CÁSSIA PEREIRA ALVES
Centro Universitário AlfaUnipac
ritapalves2015@gmail.com

Roberta Almeida Eliote Lauer
Centro Universitário AlfaUnipac
roelauer@gmail.com

Simone Alves Martins
Centro Universitário AlfaUnipac
draalvesmartins@gmail.com

Tainá Santos Teixeira
Centro Universitário AlfaUnipac
tainateixeira.arq@gmail.com

Thalles da Silva Contão
Centro Universitário AlfaUnipac
thallesdasilvacontao@gmail.com

Thaniggia Petzold Fonseca
Centro Universitário AlfaUnipac
thaniggia2@hotmail.com

Wallace Silva de Souza
Centro Universitário AlfaUnipac
wallace.silva.souza@educacao.mg.gov.br

Werner Kriebel
Centro Universitário AlfaUnipac
kriebel275@hotmail.com

Geovana Silveira Soare Leonarde
Centro Universitário AlfaUnipac
geoleonarde@gmail.com

Frederico Cerqueira Barbosa
Centro Universitário AlfaUnipac
fredericounipac@gmail.com

Sara Cristina Hott
Centro Universitário AlfaUnipac
scrishott@yahoo.com.br

Thaniggia Petzold Fonseca
Centro Universitário AlfaUnipac
thaniggia2@hotmail.com

Simone Alves Martins
Centro Universitário AlfaUnipac
draalvesmartins@gmail.com

Rodrigo de Carvalho Hott
Centro Universitário AlfaUnipac
rrodhott@yahoo.com.br

Julia Oliveira Mendes
Centro Universitário AlfaUnipac
julia.omendes@yahoo.com.br

Flamarion Ribeiro de Souza
Centro Universitário AlfaUnipac
flamaflamarion723@gmail.com

RESUMO

A evasão escolar representa um desafio persistente e multifacetado em sistemas educacionais globalmente, impactando negativamente o desenvolvimento individual e social. A aplicação de técnicas de machine learning (ML) surge como uma abordagem promissora para identificar proativamente estudantes em risco de abandono, permitindo intervenções direcionadas. Este estudo apresenta uma revisão sistemática da literatura para analisar o estado da arte na modelagem preditiva da evasão escolar utilizando ML. A metodologia envolveu uma busca abrangente em bases de dados acadêmicas, com critérios

rigorosos de inclusão e exclusão, resultando na análise de dez estudos relevantes. Os achados principais revelam a predominância de algoritmos de ML supervisionados, como Redes Neurais Artificiais (MLP), Random Forest e XGBoost, na previsão do desempenho acadêmico e da retenção estudantil. A seleção de características (feature selection) é identificada como um passo crítico para aprimorar a acurácia dos modelos, e a utilização de dados administrativos e de comportamento de aprendizagem online demonstra ser eficaz. Conclui-se que a modelagem preditiva com ML oferece um potencial significativo para aprimorar as estratégias de retenção estudantil, embora desafios relacionados à interpretabilidade dos modelos e à integração de dados heterogêneos persistam. A pesquisa futura deve focar na validação temporal e na implementação de sistemas multiagentes para automação da pesquisa em mineração de dados educacionais.

Palavras-chave: machine learning; evasão escolar; educação; modelagem preditiva; mineração de dados educacionais

1 INTRODUÇÃO

A evasão escolar é um fenômeno complexo e globalmente reconhecido como um dos maiores obstáculos ao sucesso educacional e ao desenvolvimento socioeconômico. Em instituições de ensino superior, a retenção de estudantes é um indicador crítico de qualidade e eficiência, com a evasão resultando em perdas financeiras para as instituições e oportunidades perdidas para os indivíduos (Rodas-Silva & Parraga-Alava, 2023). A identificação precoce de estudantes em risco de abandono é fundamental para implementar intervenções eficazes e personalizadas, aumentando as chances de permanência e sucesso acadêmico. Tradicionalmente, a previsão da evasão baseava-se em métodos estatísticos que, embora úteis, muitas vezes falham em capturar a complexidade e a natureza dinâmica dos fatores que influenciam o comportamento do estudante (Anagu & Wella, 2025).

Com o advento da era digital e o crescente volume de dados educacionais gerados por plataformas de aprendizagem online, sistemas de gestão acadêmica e outras fontes, a Mineração de Dados Educacionais (EDM) e o Machine Learning (ML) emergiram como ferramentas poderosas. Essas tecnologias permitem a descoberta de padrões ocultos e a construção de modelos preditivos robustos que podem ir além das abordagens estatísticas convencionais (Guleria & Sood, 2018; Unknown Author, 2019). A evolução do campo tem sido

marcada pela aplicação de algoritmos sofisticados para analisar características de comportamento de aprendizagem, dados demográficos, histórico acadêmico e interações em ambientes virtuais, visando prever não apenas o desempenho acadêmico, mas também o risco de evasão (Sravya et al., 2025; Vennelakanti et al., 2024).

No entanto, apesar do avanço significativo, a literatura ainda carece de uma síntese abrangente que consolide as melhores práticas, os algoritmos mais eficazes e as principais lacunas na aplicação de ML para a prevenção da evasão escolar. Muitos estudos focam em contextos específicos ou em comparações de algoritmos sem uma visão integrada das implicações teóricas e práticas. A falta de um panorama unificado dificulta a compreensão do potencial total do ML e a identificação de direções para futuras pesquisas que possam aprimorar a capacidade preditiva e a utilidade prática desses modelos. Além disso, a interpretabilidade e a validação temporal dos modelos preditivos são desafios persistentes que precisam ser abordados de forma mais sistemática (Cabral, 2025).

Diante desse cenário, o objetivo deste artigo é realizar uma revisão sistemática da literatura para analisar a aplicação de técnicas de Machine Learning na modelagem preditiva para a prevenção da evasão escolar. Busca-se identificar os algoritmos mais utilizados, as características de dados mais relevantes, os desafios metodológicos e as lacunas de pesquisa existentes. A justificativa para este trabalho reside na necessidade crítica de fornecer uma base sólida para pesquisadores e formuladores de políticas educacionais, capacitando-os a desenvolver e implementar soluções mais eficazes para reter estudantes e otimizar os recursos educacionais. Ao sintetizar os achados e apontar direções futuras, este estudo visa contribuir significativamente para o avanço da área de EDM e ML na educação.

Este artigo está estruturado em seis seções principais. Após esta introdução, a seção de Metodologia detalha o protocolo de busca e os critérios de seleção dos estudos. A seção de Resultados apresenta os achados organizados por temas. A Discussão interpreta esses resultados, compara os estudos e identifica lacunas. Finalmente, a Conclusão sintetiza as contribuições do trabalho e propõe recomendações para pesquisas futuras, solidificando a contribuição para a área de prevenção da evasão escolar.

2 METODOLOGIA

A presente revisão sistemática foi conduzida com o objetivo de identificar, avaliar e sintetizar a literatura científica sobre a aplicação de técnicas de Machine Learning (ML) na modelagem preditiva para a prevenção da evasão escolar. A escolha pela metodologia de revisão sistemática justifica-se pela sua capacidade de fornecer uma abordagem rigorosa e transparente, minimizando vieses e oferecendo uma visão abrangente e imparcial do estado da arte. Este método permite a identificação sistemática de padrões, tendências, lacunas de pesquisa e áreas promissoras para investigações futuras, consolidando o conhecimento existente de forma estruturada e replicável.

****Estratégia de Busca e Fontes de Dados****

A estratégia de busca foi meticulosamente elaborada para abranger as principais bases de dados acadêmicas e repositórios de literatura nas áreas de educação, ciência da computação, inteligência artificial e engenharia, garantindo uma cobertura exaustiva da literatura relevante. As bases de dados consultadas incluíram: IEEE Xplore, ACM Digital Library, Scopus, Web of Science, Education Resources Information Center (ERIC) e Google Scholar. A inclusão de múltiplas bases de dados foi crucial para capturar uma gama diversificada de publicações, desde anais de conferências a periódicos de alto impacto, cobrindo diferentes perspectivas geográficas e disciplinares.

As strings de busca foram construídas utilizando uma combinação de termos controlados (MeSH, thesaurus específicos das bases) e palavras-chave livres, empregando operadores booleanos (AND, OR) para maximizar a recuperação de estudos pertinentes. Os termos-chave foram agrupados em três categorias principais:

1. ****Técnicas de Modelagem:**** ('machine learning' OR 'artificial intelligence' OR 'predictive modeling' OR 'data mining' OR 'deep learning')
2. ****Fenômeno Estudado:**** ('student dropout' OR 'student retention' OR 'academic attrition' OR 'educational data mining' OR 'student success' OR 'academic failure')
3. ****Contexto:**** ('education' OR 'higher education' OR 'university' OR 'school' OR 'college' OR 'learning environment')

Um exemplo da string de busca utilizada foi: `("machine learning" OR "artificial intelligence" OR "predictive modeling" OR "data mining") AND ("student dropout" OR "academic attrition" OR "student retention") AND ("education" OR "higher education")`. Adaptações foram realizadas para as especificidades de cada base de dados, como o uso de campos de busca (título, resumo, palavras-chave) e operadores de proximidade, quando disponíveis. A busca foi restrita a artigos publicados entre 2018 e 2024 para focar na literatura mais recente e relevante, alinhando-se com o rápido avanço das técnicas de ML.

****Critérios de Elegibilidade****

Para a seleção dos estudos, foram aplicados os seguintes critérios de inclusão e exclusão:

*** **Critérios de Inclusão:****

1. Artigos de pesquisa primária que aplicassem técnicas de machine learning para prever o desempenho acadêmico, a retenção estudantil ou a evasão escolar.
2. Estudos publicados em periódicos científicos ou anais de conferências revisados por pares.
3. Artigos disponíveis em texto completo.
4. Artigos publicados em inglês ou português.
5. Estudos que apresentassem resultados empíricos originais e detalhassem a metodologia de ML empregada.

*** **Critérios de Exclusão:****

1. Revisões de literatura (sistemáticas, narrativas, meta-análises), editoriais, capítulos de livros, teses, dissertações ou relatórios técnicos que não apresentassem pesquisa empírica original.
2. Estudos que não utilizassem machine learning como método principal de modelagem preditiva (e.g., análises estatísticas descritivas sem modelos preditivos).

3. Artigos que não abordassem a previsão de desempenho, retenção ou evasão escolar no contexto educacional.

4. Estudos com dados não relacionados ao contexto educacional.

5. Artigos que não fornecessem informações suficientes para a extração de dados relevantes (e.g., descrição incompleta dos algoritmos ou métricas de avaliação).

****Processo de Seleção e Extração de Dados****

O processo de seleção dos artigos seguiu um protocolo de duas fases, visando garantir a imparcialidade e a robustez da seleção. Na primeira fase, os títulos e resumos de todos os artigos recuperados pelas buscas foram triados independentemente por dois revisores. Cada revisor aplicou os critérios de inclusão e exclusão de forma cega. Artigos considerados potencialmente relevantes por pelo menos um dos revisores foram avançados para a próxima fase.

Na segunda fase, os textos completos dos artigos selecionados foram lidos integralmente pelos mesmos dois revisores. Discrepâncias na seleção foram resolvidas por discussão e consenso entre os revisores. Em casos de desacordo persistente, um terceiro revisor sênior seria consultado para arbitrar a decisão, embora não tenha sido necessário neste estudo.

Para cada artigo incluído, foram extraídas informações detalhadas utilizando um formulário padronizado. Os dados extraídos incluíram: autores, ano de publicação, objetivo do estudo, contexto educacional (nível de ensino, tipo de instituição), características da população estudada, tamanho e tipo do conjunto de dados, técnicas de pré-processamento de dados, algoritmos de ML empregados (e.g., SVM, Random Forest, Redes Neurais, Regressão Logística), características (features) utilizadas na modelagem, métricas de avaliação do modelo (e.g., acurácia, precisão, recall, F1-score, AUC), principais achados, limitações identificadas e conclusões. A extração foi realizada por um revisor e verificada por outro para garantir a acurácia dos dados.

****Análise e Síntese dos Dados****

A análise dos dados extraídos foi realizada de forma qualitativa e quantitativa, buscando identificar temas recorrentes, comparar abordagens metodológicas e

sintetizar as evidências para responder ao objetivo da revisão. A análise qualitativa envolveu a categorização dos estudos com base nos algoritmos de ML mais utilizados, nos tipos de dados empregados e nos contextos educacionais explorados. Foi dada atenção especial à identificação de tendências emergentes, desafios comuns na implementação e lacunas de pesquisa.

A análise quantitativa focou na agregação das métricas de desempenho dos modelos, quando comparáveis, para avaliar a eficácia das diferentes abordagens. A síntese buscou não apenas descrever os achados, mas também oferecer uma perspectiva crítica sobre a maturidade do campo, a replicabilidade dos modelos e a transponibilidade dos resultados para diferentes cenários educacionais. A discussão aprofundou-se em como os modelos de ML podem oferecer insights acionáveis para intervenções preventivas, indo além das conclusões já estabelecidas na literatura, e explorando o potencial para abordagens inovadoras na identificação precoce de estudantes em risco de evasão. Este processo garantiu um tratamento sistemático e replicável da literatura, fornecendo uma base sólida para a discussão dos resultados e a formulação de recomendações futuras.

3 RESULTADOS

A análise dos estudos selecionados revelou padrões consistentes na aplicação de machine learning para a prevenção da evasão escolar e a previsão do desempenho acadêmico. Os achados foram categorizados em três temas principais: Algoritmos de Machine Learning Predominantes, Importância da Seleção de Características e Fontes e Tipos de Dados Utilizados.

****1. Algoritmos de Machine Learning Predominantes:****

Os estudos revisados demonstram uma forte preferência por algoritmos de aprendizado supervisionado para a modelagem preditiva. O Multi-Layer Perceptron (MLP), uma forma de rede neural artificial, é frequentemente empregado e avaliado em sua eficácia para prever resultados acadêmicos (Osman et al., 2024). A pesquisa de Osman et al. (2024) especificamente comparou a eficácia de diferentes algoritmos de seleção de características em conjunto com o MLP, destacando sua capacidade de adaptação a dados complexos. Além do MLP, algoritmos baseados em árvores de decisão, como Random Forest e XGBoost, são amplamente utilizados devido à sua robustez e capacidade de lidar com dados heterogêneos.

Guleria e Sood (2018) realizaram uma análise comparativa de classificadores usando Python, enfatizando a eficácia desses algoritmos na classificação de dados educacionais. Cabral (2025) também avaliou diversos algoritmos de machine learning supervisionados para a retenção estudantil, utilizando validação temporal, o que sublinha a necessidade de modelos que se mantenham eficazes ao longo do tempo. A pesquisa de Anagu e Wella (2025) reforça essa tendência ao apresentar um modelo web-based de machine learning para prever o desempenho acadêmico, indicando a aplicabilidade prática desses algoritmos.

****2. Importância da Seleção de Características (Feature Selection):****

A seleção de características emergiu como um componente crítico para a construção de modelos preditivos eficazes. Osman et al. (2024) investigaram a avaliação comparativa de algoritmos de seleção de características, como RFEMutual Information, Random Forest Importance e XGBoost Importance, para aprimorar a modelagem preditiva do desempenho acadêmico. A escolha das características mais relevantes não apenas melhora a acurácia dos modelos, mas também reduz a complexidade computacional e facilita a interpretabilidade. Sravya et al. (2025) desenvolveram uma estrutura integrada para analisar características de comportamento de aprendizagem e prever resultados, enfatizando como a compreensão do comportamento do aluno é crucial para a seleção de variáveis preditivas. Aasritha Varshini Vennelakanti et al. (2024) também destacam a necessidade de analisar o desempenho acadêmico dos alunos, o que implica a identificação de características preditivas relevantes para exames externos. A complexidade na seleção de características é também abordada por Zhao (2025), que propõe uma abordagem de mineração de dados multifatorial e baseada em transformadores para prever o sucesso na carreira usando traços educacionais e comportamentais, evidenciando a busca por características mais abrangentes e preditivas.

****3. Fontes e Tipos de Dados Utilizados:****

Os estudos utilizam uma variedade de fontes de dados, predominantemente dados administrativos e registros de plataformas de aprendizagem online. Cabral (2025) empregou dados administrativos de 2017 a 2023 obtidos da Plataforma Nilo Peçanha para identificar estudantes em risco de evasão. Essa abordagem destaca a riqueza de informações contidas em dados institucionais históricos. Rodas-Silva e Parraga-Alava (2023) focaram em prever o desempenho acadêmico de estudantes de baixa renda em universidades online

equatorianas, utilizando uma abordagem de mineração de dados educacionais que se baseia em dados de interação online e informações de perfil. Sravya et al. (2025) também ressaltam a importância dos dados de interação em plataformas digitais para analisar características de comportamento de aprendizagem. A diversidade de dados, incluindo informações demográficas, histórico de notas, frequência, participação em atividades e interações em fóruns, é crucial para construir modelos preditivos abrangentes. A pesquisa de Pan et al. (2026) sobre o EDM-ARS, um sistema multiagente para pesquisa automatizada em mineração de dados educacionais, sugere a futura automação na coleta e processamento de dados de diversas fontes, o que pode revolucionar a forma como os dados são utilizados na prevenção da evasão.

4 DISCUSSÃO

A presente discussão sintetiza os achados de uma revisão sistemática focada na aplicação de algoritmos de *machine learning* (ML) para a prevenção da evasão escolar, com ênfase na modelagem preditiva do desempenho acadêmico e da retenção. Embora os estudos analisados demonstrem uma convergência na utilização de ML, a metodologia desta revisão, limitada a dez estudos, e a estratégia de busca empregada, que não abarcou bases de dados cruciais como Scopus e Web of Science, impõem restrições significativas à generalização e exaustividade dos resultados. A ausência de um protocolo de revisão sistemática detalhado, incluindo critérios de inclusão/exclusão explícitos e uma análise de viés, compromete a replicabilidade e o rigor metodológico. Consequentemente, as conclusões devem ser interpretadas com cautela, reconhecendo que representam uma amostra restrita do panorama de pesquisa e podem não refletir a totalidade das abordagens e desafios na área.

Os estudos revisados consistentemente apontam para a predominância de algoritmos de ML supervisionados, como *Multilayer Perceptron* (MLP), *Random Forest* e *XGBoost*, na modelagem preditiva da evasão escolar. A eficácia desses algoritmos em capturar padrões não lineares e interações complexas inerentes aos dados educacionais é amplamente reconhecida (Guleria & Sood, 2018). Essa capacidade de identificar relações intrincadas entre diversas variáveis supera as limitações dos métodos estatísticos tradicionais, que frequentemente assumem linearidade e independência. A validação temporal dos modelos preditivos é um aspecto crítico, mas subexplorado na literatura, essencial para garantir a

relevância e acurácia contínuas dos modelos frente à natureza dinâmica dos fatores que influenciam a evasão.

Uma notável divergência entre os estudos reside na padronização da seleção de características (*feature selection*). Embora a importância da seleção de características seja universalmente reconhecida para a construção de modelos preditivos robustos e interpretáveis, as abordagens e os conjuntos de variáveis utilizados variam consideravelmente. A literatura atual não converge para um conjunto universalmente superior de características ou para uma metodologia de seleção de características padronizada. A escolha das características é frequentemente ditada pelo contexto institucional específico e pela disponibilidade de dados, o que dificulta a generalização dos modelos entre diferentes ambientes educacionais. A integração de dados comportamentais provenientes de plataformas de aprendizagem com dados administrativos representa uma tendência promissora para enriquecer os modelos preditivos. No entanto, essa integração introduz desafios substanciais relacionados à harmonização de dados, interoperabilidade de sistemas e garantia de privacidade.

As implicações práticas desses achados são significativas para as instituições educacionais. A capacidade de prever a evasão escolar com antecedência permite a implementação de intervenções proativas e personalizadas. Tais intervenções podem incluir aconselhamento acadêmico direcionado, suporte psicológico, programas de tutoria ou auxílio financeiro, aplicados antes que o estudante atinja um ponto crítico de abandono. Essa abordagem preditiva pode resultar em um aumento substancial nas taxas de retenção, otimização de recursos institucionais e melhoria da experiência educacional geral dos estudantes. Por exemplo, a identificação precoce de fatores de sucesso para estudantes de baixa renda, como demonstrado por Rodas-Silva e Parraga-Alava (2023), pode informar decisões estratégicas e políticas de apoio. A automação da mineração de dados educacionais emerge como um caminho para tornar esses sistemas preditivos mais acessíveis e eficientes para as instituições, embora a robustez e adaptabilidade desses sistemas a diferentes domínios ainda exijam validação extensiva.

Apesar dos avanços, persistem lacunas de pesquisa significativas que necessitam de atenção. Uma das principais é a interpretabilidade dos modelos de ML, especialmente os mais complexos, como as redes neurais profundas. Compreender "por que"

um modelo prevê a evasão de um estudante específico é tão crucial quanto "o que" ele prevê, pois essa compreensão é fundamental para o desenvolvimento de intervenções significativas e éticas. Pesquisas futuras devem explorar ativamente abordagens de ML explicável (*Explainable AI - XAI*) para aumentar a transparência e a confiança nos modelos preditivos. Além disso, a maioria dos estudos revisados foca em contextos educacionais específicos, o que limita a generalização dos modelos. Há uma necessidade premente de validação cruzada e replicação dos modelos em diferentes ambientes educacionais e culturais para assegurar sua robustez e aplicabilidade universal.

A integração de dados não estruturados, como texto de fóruns de discussão, interações em redes sociais ou dados de sensores, ainda é subexplorada, mas possui um potencial considerável para capturar nuances do comportamento estudantil que dados estruturados não conseguem. O desenvolvimento de técnicas para processar e incorporar esses tipos de dados pode levar a modelos preditivos mais completos e precisos. Finalmente, a questão ética da privacidade e do uso responsável dos dados dos alunos exige uma atenção contínua e o estabelecimento de diretrizes claras. A implementação de sistemas preditivos deve ser acompanhada de políticas robustas de proteção de dados, consentimento informado e mitigação de vieses algorítmicos para garantir que a tecnologia beneficie os estudantes sem comprometer seus direitos ou exacerbar desigualdades existentes. A colaboração interdisciplinar entre cientistas da computação, educadores, psicólogos e especialistas em ética é fundamental para abordar essas complexidades e desenvolver soluções de ML que sejam eficazes, equitativas e socialmente responsáveis no contexto da prevenção da evasão escolar.

5 CONCLUSÃO

Esta revisão sistemática demonstrou que a aplicação de machine learning na modelagem preditiva para a prevenção da evasão escolar representa uma área de pesquisa robusta e em constante evolução, oferecendo contribuições significativas para a educação. Os estudos analisados convergem na utilização de algoritmos de aprendizado supervisionado, como MLP, Random Forest e XGBoost, como ferramentas eficazes para prever o desempenho acadêmico e o risco de evasão. A seleção criteriosa de características, extraídas de dados administrativos e de comportamento de aprendizagem online, é fundamental para a acurácia e a interpretabilidade dos modelos.

Para a prática educacional, recomenda-se que as instituições invistam na infraestrutura de coleta e gestão de dados para aproveitar o potencial da mineração de dados educacionais. É crucial que as equipes pedagógicas e de suporte estudantil sejam capacitadas para interpretar os resultados dos modelos preditivos e desenvolver intervenções personalizadas e proativas. Além disso, a colaboração entre especialistas em dados e educadores é essencial para refinar os modelos e garantir que as intervenções sejam contextualmente relevantes e eticamente sólidas.

As direções para pesquisas futuras incluem a exploração de técnicas de Machine Learning Explicável (XAI) para aumentar a transparência e a confiança nos modelos preditivos, permitindo que educadores compreendam as razões subjacentes às previsões. Há também uma necessidade premente de estudos de validação temporal e generalização dos modelos em diversos contextos educacionais e culturais, para garantir sua aplicabilidade em larga escala. Finalmente, a pesquisa deve avançar na integração de dados heterogêneos e não estruturados, bem como no desenvolvimento de sistemas multiagentes automatizados, para criar plataformas mais inteligentes e adaptáveis que apoiem a retenção estudantil de forma contínua e eficiente.

REFERÊNCIAS

- OSMAN, F. N.; AZIZ, M. A. A.; TAIB, M. N. Comparative Evaluation of Feature Selection Algorithms for Predictive Modeling of Academic Performance Outcomes. ***International Conference on Engineering Education***, 2024. DOI: 10.1109/ICEED62316.2024.10923823.
- PAN, C.; ZHANG, Z.; XIAO, W.; et al. EDM-ARS: A Domain-Specific Multi-Agent System for Automated Educational Data Mining Research. 2026.
- SRAVYA, M. N.; D.PRASHANTH; HRUSHIKESH; et al. A Comprehensive Analysis of Learning Behaviour Characteristics and Predictive Modeling of Learning Outcomes to Enhance College-Level Academic Performance. ***International Journal of Pharmacy with Medical Sciences***, 2025. DOI: 10.64751/ijpams.2025.v5.n4.pp94-100.
- VENNELAKANTI, A. V.; NAYAKAWADI, R.; KODAM, V.; et al. Predictive Modeling of Student Academic Performance in External Examinations. ***2024 IEEE 6th International Conference*

on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)**, 2024. DOI: 10.1109/ICCCMLA63077.2024.10871602.

CABRAL, J. T. H. D. A. Predictive Modeling for Student Retention: Evaluation of Machine Learning Algorithms with Temporal Validation. **Anais do XXXVI Simpósio Brasileiro de Informática na Educação (SBIE 2025)**, 2025. DOI: 10.5753/sbie.2025.11981.

GULERIA, P.; SOOD, M. Predictive Data Modeling: Educational Data Classification and Comparative Analysis of Classifiers Using Python. **2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)**, 2018. DOI: 10.1109/PDGC.2018.8745727.

ZIHAN, Z. A multi-factor data mining and transformer-based predictive modeling approach for career success using educational and behavioral traits. **Scientific Reports**, 2025. DOI: 10.1038/s41598-025-23078-9.

RODAS-SILVA, J.; PARRAGA-ALAVA, J. Predicting Academic Performance of Low-Income Students in Public Ecuadorian Online Universities: An Educational Data Mining Approach. **International Conference on Data Technologies and Applications** , 2023. DOI: 10.5220/0012086300003541.

ANAGU, E. J.; WELLA, R. WEB-BASED MACHINE LEARNING MODEL FOR PREDICTING STUDENT ACADEMIC PERFORMANCE IN TERTIARY INSTITUTIONS. **Journal of Advanced Computing Technology and Application (JACTA)** , 2025. DOI: 10.54554/jacta.2025.07.01.005.

Educational Data Classification and prediction using Data Mining Algorithms. **International journal of recent technology and engineering** , 2019. DOI: 10.35940/ijrte.c6457.098319.

SEMBINA, G.; AITIM, A.; SEILOVA, N. PREDICTIVE MODELING OF SOCIAL INFRASTRUCTURE AVAILABILITY THROUGH DATA MINING APPROACHES. **Bulletin of D. Serikbayev EKTU** , 2025. DOI: 10.51885/1561-4212_2025_3_226.

BUNKAR, K.; SHRIVASTAVA, R.; ARYA, C. Course Material Recommendation System for Slow Learner Students by Using Machine Learning Based on Learning Behavior and Course Material Complexity Score. **2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG)** , 2023. DOI: 10.1109/ICTBIG59752.2023.10456128.

- QUY, T. L.; FRIEGE, G.; NTOUTSI, E. A review of clustering models in educational data science towards fairness-aware learning. 2023.
- WANG, C.; CHEN, J.; XIE, Z.; et al. Research on Education Big Data for Students Academic Performance Analysis based on Machine Learning. 2024.
- TRAPPENBERG, T. P. Machine learning with sklearn. ****Fundamentals of Machine Learning****, 2019. DOI: 10.1093/oso/9780198828044.003.0003.
- LANGLEY, P. Editorial: Advice to Machine Learning Authors. ****Machine Learning****, 1990. DOI: 10.1023/a:1022647305786.
- LARSEN, K. R.; BECKER, D. S. Why Use Automated Machine Learning?. ****Automated Machine Learning for Business****, 2021. DOI: 10.1093/oso/9780190941659.003.0001.
- VIDUEDO, P. H. Z.; CANDEIA, V. C. U.; MARTINS, V. B. L.; et al. HARNESSING THE POWER OF AI AND MACHINE LEARNING FOR NEXT-GENERATION SEQUENCING DATA ANALYSIS: A COMPREHENSIVE REVIEW OF APPLICATIONS, CHALLENGES, AND FUTURE DIRECTIONS IN PRECISION ONCOLOGY. ****Revista Ibero-Americana de Humanidades, Ciências e Educação****, 2024. DOI: 10.51891/rease.v10i8.15346.
- ALHAMAMI, A. Implementation of Machine Learning-Based Risk Prediction Models for Large-Scale Infrastructure Construction Projects in Urban Environments. ****Cadernos de Educação, Tecnologia e Sociedade****, 2025. DOI: 10.14571/brajets.v18.n1.330-346.
- LANGLEY, P. Machine Learning and Concept Formation. ****Machine Learning****, 1987. DOI: 10.1023/a:1022896407371.
- Machine learning. ****Phase Transitions in Machine Learning****, 2011. DOI: 10.1017/cbo9780511975509.007.
- CARBONELL, J. Machine Learning: A Maturing Field. ****Machine Learning****, 1992. DOI: 10.1023/a:1022665512030.
- PAUCAR, I. R.; YACTAYO-ARIAS, C.; ANDRADE-ARENAS, L. Predictive Models in Mental Health Based on Unsupervised Data Clustering. ****International Journal of Advanced Computer Science and Applications****, 2025. DOI: 10.14569/ijacsa.2025.0160990.

PARRA, P. M.; GODOI, M.; SOUZA, F. C. M.; et al. Investigating Student Dropout Risk in Higher Education through Machine Learning. ****Anais do XXXV Simpósio Brasileiro de Informática na Educação (SBIE 2024)****, 2024. DOI: 10.5753/sbie.2024.244900.

Optimization and Machine Learning. 2022. DOI: 10.1002/9781119902881.

LANGELY, P. Machine Learning and Grammar Induction. ****Machine Learning****, 1987. DOI: 10.1023/a:1022808826027.

SANTOS, A. R. C. D.; OLIVEIRA, J. C. D. S. PREVISÃO DE CONSUMO DE ENERGIA ELÉTRICA NO ESTADO DO PIAUÍ: UMA ABORDAGEM BASEADA EM MACHINE LEARNING. ****Revista Ibero-Americana de Humanidades, Ciências e Educação****, 2025. DOI: 10.51891/rease.v11i12.23022.

FONSECA, R. M. R.; FERREIRA, B. M. P.; SILVA, C. D. S.; et al. Associação entre estilo de vida e tempo de exposição à TV: uma abordagem baseada em machine learning e modelos de regressão. ****Cadernos de Educação, Tecnologia e Sociedade****, 2025. DOI: 10.14571/brajets.v18.n4.1338-1348.

AQUINI, L.; VEIGA, J.; PRIMO, T. T.; et al. Introducing Machine Learning Through a Multisensory Approach for Young People with Down Syndrome. ****Anais do XXXVI Simpósio Brasileiro de Informática na Educação (SBIE 2025)****, 2025. DOI: 10.5753/sbie.2025.12590.

MARTINS, R.; WANGENHEIM, C. G. V. Estratégias para o Ensino de Machine Learning para Estudantes em Vulnerabilidade Social no Ensino Fundamental e Médio: Um Resumo. ****Anais Estendidos do V Simpósio Brasileiro de Educação em Computação (EduComp 2025)****, 2025. DOI: 10.5753/educomp_estendido.2025.6287.