

ANÁLISE DO DESEMPENHO DAS ESCOLAS DO ESTADO DE SÃO PAULO NA PRÉ-PANDEMIA E PANDEMIA

Paulo Henrique ¹

Maria das Graças J. M. Tomazela ²

Rosana Helena Nunes ³

RESUMO

A pandemia de COVID-19 trouxe impactos significativos para diversos setores da sociedade, incluindo a educação. No estado de São Paulo, as escolas enfrentaram mudanças drásticas durante a pandemia, uma vez que tiveram que se adequar rapidamente ao ensino remoto, o que implicou em investimentos em tecnologia e capacitação de professores. Assim, este trabalho teve como objetivo analisar a situação educacional antes e durante a pandemia no estado de São Paulo. Com a utilização das técnicas de Aprendizado de Máquina supervisionado, o estudo se ancorou em práticas apresentadas em trabalhos relacionados. A pesquisa, de natureza experimental, teve início com a coleta de dados do IDESP (Índice de Desenvolvimento da Educação do Estado de São Paulo). Após a coleta, os dados foram limpos e transformados, e em seguida, submetidos à clusterização utilizando Jupyter Notebook. Gráficos elaborados no Excel ajudaram na análise dos clusters. A comparação dos clusters pré-pandemia e durante a pandemia revelou mudanças importantes na distribuição demográfica e no desempenho acadêmico dos alunos. Antes da pandemia, a maioria dos alunos estava nos Anos Iniciais, seguidos pelos Anos Finais e pelo Ensino Médio. Observou-se que o desempenho acadêmico era melhor em escolas com menos alunos por unidade escolar. Já durante a pandemia, houve um aumento na retenção por frequência e rendimento insuficiente em todos os níveis escolares, destacando os desafios da adaptação ao ensino remoto. A pandemia intensificou esses desafios, especialmente, para alunos dos Anos Iniciais e Finais. Também foi observada uma mudança na distribuição étnica, com uma maior concentração de alunos brancos e pardos. Estes dados sublinham a necessidade de uma abordagem educacional adaptativa, considerando as características demográficas e as características acadêmicas específicas. A pesquisa ressalta a importância da colaboração entre escolas, comunidades e autoridades educacionais para implementar estratégias eficazes, garantindo educação de qualidade para os alunos.

Palavras-chave: Inteligência artificial, Mineração de dados, Desempenho escolar.

INTRODUÇÃO

Segundo o Banco Mundial (2021), a crise global de Covid-19 ameaçou esforços de duas décadas na educação. Em meio à pandemia, apesar dos governos disponibilizarem pacotes de estímulo fiscal para poderem injetar mais recursos no setor educacional, reve-

¹ Graduado do Curso de Análise e Desenvolvimento de Sistemas da Faculdade de Tecnologia de Sorocaba, FATEC Sorocaba, paulofernandaraitz@gmail.com;

² Doutora pelo Curso de Engenharia de Produção da Universidade Metodista de Piracicaba, UNIMEP, graca.tomazela@fatec.sp.gov.br;

³ Doutora pelo Curso de Língua Portuguesa da Pontifícia Universidade Católica – PUC-SP, rosana.nunes@fatec.sp.gov.br.



lou-se que 65% dos países de renda baixa reduziram seu orçamento na educação, comparado com 35% em nações ricas. Além disso, os pais mais pobres aplicaram apenas 2% dos recursos disponibilizados na educação.

O Brasil passou por momentos de altos e baixos, politicamente, que se destacaram por atrasar muito o processo de instalação das aulas virtuais, e distribuição de recursos, afetando o combate na defasagem da aprendizagem e manutenção das metas anuais determinadas para escolas públicas da rede estadual.

O indicador do IDESP busca avaliar a qualidade das escolas da rede estadual do estado de São Paulo e prever condições favoráveis de melhoria na qualidade de seu desempenho, estabelecendo metas anuais sendo composta por critérios avaliativos em exames de proficiência do SARESP (o quanto os alunos aprenderam) e o fluxo escolar (quanto tempo demandou essa aprendizagem), os critérios são contínuos permitindo acompanhar sua evolução anualmente e a qualidade do ensino oferecido pelas escolas das redes públicas.

O momento da pandemia representou uma ruptura nos costumes, forma de pensar e principalmente de questionar verdades absolutas da sociedade. Entre os seus questionamentos estão a educação pública e especialmente o desempenho escolar.

Considerando esse grande volume de dados e a necessidade de avaliar esses dados, o uso das tecnologias de informação pode contribuir para melhorar a eficiência da gestão desses dados educacionais. Nesse sentido, as técnicas de mineração de dados permitem a identificação de relacionamentos implícitos em grandes bancos de dados que envolvam muitas variáveis. Com isso, é possível descobrir novos padrões, dar maior precisão em padrões conhecidos e modelar fenômenos do mundo real (Han, Kamber e Pei, 2012).

Assim, o objetivo deste trabalho foi dar suporte aos processos de definição de políticas públicas educacionais, a partir da utilização de técnicas de mineração de dados.

METODOLOGIA

A metodologia utilizada foi de caráter experimental na qual, segundo Gil (2008), determina-se um objeto de estudo, selecionam-se as variáveis que podem influenciá-lo, e definem-se as formas de controlar e observar os efeitos que essas variáveis produzem no objeto.

Para o desenvolvimento deste trabalho, foi utilizado o processo de KDD. Na etapa de pré-processamento, inicialmente foi realizada a extração dos dados da relação completa de alunos matriculados nos anos de 2017 a 2022. Os dados foram obtidos no site do



IDESP os arquivos são divididos por ano e foram unificados em dois uma chamada tabela pré-pandemia com os anos de 2017 a 2019 e o outro tabela pandemia com os anos de 2020 a 2022. Em seguida, ainda utilizando o Jupyter Notebook, foi realizado o tratamento de dados faltantes e de valores outliers a fim de se obter uma base de dados de qualidade.

Após a etapa de pré-processamento, foi iniciada a etapa de mineração de dados e, para isso, foi utilizado o algoritmo k-means, a fim de realizar a clusterização dos dados. Após a realização de diversos experimentos com diferentes números de clusters, foi definido 3 clusters como valor ideal. Todo o processo foi realizado no ambiente Jupyter Notebook.

Na última etapa, o pós-processamento, realizou-se a análise dos clusters gerados, os resultados obtidos foram documentados e a pesquisa foi finalizada.

Processo de pré-processamento de dados

Inicialmente foi realizada a extração dos dados da relação completa de alunos matriculados nos anos de 2017 a 2022 por meio do site de dados abertos IDESP. As bases de dados estavam divididas por ano, ficando cada uma em um arquivo diferente no (formato .csv). Assim, realizou-se a consolidação dessas bases em dois arquivos por meio do Jupyter Notebook e da biblioteca Pandas (linguagem de programação Python). Como o estudo é sobre o desempenho das escolas do Estado de São Paulo, foram excluídas colunas desnecessárias para os estudos, dessa forma, das 53 colunas iniciais, foram selecionados 7 atributos.

Diante da necessidade de lidar com a quantidade significativa de dados ausentes, optou-se pela criação de novos atributos como parte de uma investigação mais detalhada. Assim, foram criados diversos campos, calculados a partir de médias ou somatórias

Depois da criação de todos esses novos atributos, foram excluídos os julgados desnecessários e ficaram em definitivo 24 atributos descritos a seguir:

- Gênero: Foram criadas colunas "Feminino" e "Masculino" para contabilizar os alunos de cada sexo por escola.
- Cor/Raça: Criadas colunas para identificar alunos classificados como "BRANCOS", "PRETOS", "PARDOS", "AMARELOS", "INDÍGENAS" e "NÃO DECLARADOS".
- Rendimento: Adicionadas categorias para rendimento escolar: "Aprovado", "Progressão Parcial", "Retido por Frequência Insuficiente" e "Retido por Rendimento Insuficiente".



- Séries: Criados atributos para as séries escolares (de "SERIE 1" a "SERIE 12") e agrupados em "Anos Iniciais" (1º ao 5º ano), "Anos Finais" (6º ao 9º ano) e "Ensino Médio" (10º ao 12º ano).
- Rendimento por Período Escolar: Novas colunas combinaram rendimento com os períodos escolares (e.g., "Aprovados Anos Iniciais").
- Médias por Escola: Calculadas médias ajustadas com base no número de alunos em cada escola.
- Períodos Escolares: Atributo "Períodos" criado para identificar combinações de níveis de ensino lecionados em cada escola (e.g., "Anos Iniciais e Finais", "Todos").

Processo de mineração de dados

Com o final da etapa do processo de pré-processamento, realizou-se a normalização dos atributos e a etapa de mineração de dados, no Jupyter Notebook. Foram introduzidas as bibliotecas Matplotlib e Scikit-learn, ambas da linguagem de programação Python para visualização gráfica dos dados e a manipulação dessas tarefas. Nessa fase, utilizou-se a técnica de clusterização e conduziram-se diversos experimentos para definir o número exato de clusters, uma vez que no algoritmo k-means o número de clusters não é predefinido.

Assim, inicialmente, empregou-se a curva de Elbow, um método usado para determinar o número ideal de clusters, com base na métrica de compactação desses clusters. Além disso, foram utilizadas 3 métricas para identificar a qualidade do agrupamento: índices Davies Bouldin, Silhouette e Calinski Harabasz.

A partir dos resultados obtidos com a curva de Elbow e os valores das métricas utilizados, definiu-se que seriam utilizados 3 clusters, para a sequência do processo de mineração de dados.

As mesmas técnicas e medidas foram utilizadas para os dados da tabela pré-pandemia. Como os resultados foram semelhantes aos dados da pandemia, o número de clusters adotado, para esse conjunto de dados, também foi 3.

A etapa de pós-processamento trata-se das análises dos clusters, que são apresentadas na seção de resultados da pesquisa.

REFERENCIAL TEÓRICO



O referencial teórico aborda o desempenho dos estudantes da educação básica no período da pandemia da Covid-19, bem como a teoria sobre o processo de mineração de dados.

Desempenho escolar

Segundo Fernandes (2008) o Brasil é parceiro internacional da International Association for the Evaluation of Educational Achievement (IEA), por meio do Ministério da Educação e Cultura (MEC), demonstrando que desde sua participação nas avaliações internacionais, ocorreu uma elevação na busca pela melhoria da qualidade oferecida nos sistemas educacionais, disponibilizando mecanismos para obtenção de dados e informações confiáveis sobre os alunos avaliados, para melhor análise do sistema educacional do país.

Segundo a Secretaria do Estado de São Paulo (São Paulo, 2011) em 2008, começou a implementação do currículo unificado na rede pública estadual, proposta curricular que emerge um diagnóstico com influência nas avaliações internacionais, contribuindo para o direcionamento de programas e projetos educacionais. O IDESP é o indicador de qualidade, escolhido pela rede estadual do Estado de São Paulo para acompanhar esse processo, por meio das avaliações de desempenho e fluxo escolar (São Paulo, 2007).

Fluxo escolar

Setubal (2000) destaca que o fluxo escolar, diferentemente do que muitos pensam ao abordar o tema, não se aplica apenas aos alunos, mas também aos professores, diretores e demais envolvidos na rotina educacional de uma determinada escola. Programas de regulamentação do fluxo escolar são implantados desde os anos 1990, buscando a diminuição de fatores de desigualdade que incluem diferenças de proficiência, alfabetização e taxa líquida de matrículas. Segundo a Secretaria do Estado de São Paulo (São Paulo, 2020), no auge da pandemia do Covid-19, o governo do estado de São Paulo tomou iniciativa própria na disponibilização de recursos para o setor da educação, visto que a solução provisória para continuação das atividades educacionais foi a implementação do uso de tecnologias digitais, ocorrendo adaptações e alterações estruturais no setor.

Proficiência

Klein (2003) afirma que proficiência é o resultado do aprendizado do aluno, indicado por meio da junção do conteúdo escolar e das metodologias aplicadas ao longo desse período. Assim, validar o entendimento de questões socioeconômicos comparativas



entre alunos próximos e distantes é fundamental na análise da proficiência no desempenho escolar.

O IDESP procura qualificar a capacidade de prever condições da qualidade de ensino das escolas, determinando metas anuais evolutivas, por meio de avaliações de fluxo escolar e proficiência, para isso disponibiliza uma grande base de dados disponíveis de ano em ano, dados esses que vão ser fruto de estudo dessa pesquisa, para identificar o desempenho dos estudantes no período pré-pandemia e pandemia.

Descoberta do conhecimento em base de dados

A informatização generalizada das mais diversas áreas, o aumento de capacidade e redução de custos dos dispositivos de armazenamento e a evolução dos dispositivos de coleta de dados, como leitores de código de barras e sensores, propiciaram o armazenamento de grandes quantidades de dados. Entretanto, a exploração desse grande volume de dados é muito difícil de ser realizada sem a utilização de técnicas e ferramentas que possibilitem a transformação dos dados em informação e conhecimento úteis.

A utilização de ferramentas de Mineração de Dados torna possível a análise dessas enormes quantias de dados (Tsai, 2013). De acordo com Tsai (2012), essa tecnologia fornece diversas metodologias para a tomada de decisão, resolução de problemas, análise, planejamento, diagnóstico, detecção, integração, prevenção, aprendizagem e inovação.

A mineração de dados faz parte de um processo maior denominado descoberta de conhecimento em base de dados ou KDD, do inglês knowledge Discovery in DataBase.

O processo de KDD é constituído de várias etapas operacionais, normalmente definidas como: pré-processamento, mineração de dados e pós-processamento (Han, Kamber e Pei, 2012).

A mineração de dados é composta por tarefas. A clusterização é uma dessas tarefas e tem como objetivo encontrar grupos (clusters) em um conjunto de dados, de modo que as observações dentro de cada grupo sejam similares entre si e diferentes das observações em outros grupos. O algoritmo K-means foi utilizado nesta pesquisa.

RESULTADOS E DISCUSSÃO

Os resultados, etapa referente ao pós-processamento, estão divididos em três partes, inicialmente apresentam-se as análises dos dados pré-pandemia, em seguida da análise dos dados da pandemia e, por último, a comparação entre os dados pré-pandemia e pandemia. Após as análises, apresentam-se as discussões dos resultados desse trabalho em relação às pesquisas relacionadas.



Análise dos clusters gerados com os dados do período pré-pandemia

O Cluster 0 apresenta uma distribuição equilibrada de gênero, refletindo a proporção nacional de alunos, com um destaque significativo para a diversidade étnica e um equilíbrio na presença de alunos brancos, pretos e pardos. As escolas têm maior incidência nos Anos Iniciais e Finais e mostram padrões distintos de aprovação e retenção. Os desafios principais se concentram na retenção por rendimento insuficiente, especialmente nos Anos Finais, onde se observa uma maior incidência de retenção.

O Cluster 1 é caracterizado por um maior número médio de alunos por escola e uma distribuição de gênero ligeiramente equilibrada. Este cluster destaca-se pela sua composição demográfica diversificada e uma alta prevalência de escolas nos Anos Iniciais. A avaliação do desempenho acadêmico revela que a média de alunos aprovados é consideravelmente alta, mas a retenção por rendimento insuficiente e a progressão parcial indicam áreas de preocupação, particularmente nos Anos Finais e no Ensino Médio. Já o Cluster 2, embora menor em número de instâncias, apresenta uma média alta de alunos por escola. A distribuição de gênero é equilibrada e a diversidade étnica é notável, com uma presença significativa de alunos brancos, pretos e pardos. As escolas neste cluster estão majoritariamente nos Anos Iniciais, e a análise do desempenho acadêmico mostra uma alta média de aprovação, mas também aponta para desafios em termos de retenção por rendimento insuficiente e frequência, especialmente no Ensino Médio.

Análise dos clusters gerados com os dados do período pandemia

A partir dos dados educacionais da pandemia, o Cluster 0 foi caracterizado por uma grande população escolar, com uma distribuição equilibrada entre gêneros. Apresenta uma alta proporção de alunos brancos e pardos, refletindo uma diversidade étnica significativa. As escolas deste cluster concentram-se predominantemente nos Anos Iniciais e Finais, enfrentando desafios expressivos relacionados à retenção por rendimento insuficiente, especialmente nos Anos Finais. Além disso, a distribuição de alunos aprovados, retenções por rendimento insuficiente, progressão parcial e retenções por frequência varia significativamente entre os diferentes níveis de ensino.

Já o Cluster 1 agrupa escolas com uma média maior de alunos e uma distribuição de gênero quase equilibrada. Assim como no Cluster 0, há uma considerável diversidade étnica, com maior concentração nos Anos Iniciais. Os desafios principais também estão associados à retenção por rendimento insuficiente, embora as médias de aprovação e retenção por frequência apresentem variações expressivas entre os níveis de ensino.



Por fim, o Cluster 2 abrange escolas com uma média menor de alunos, mantendo um equilíbrio quase perfeito na distribuição de gênero. Assim como os outros clusters, exibe diversidade étnica, com predominância de alunos nos Anos Iniciais. Este cluster se destaca por ter a menor incidência de progressão parcial e retenções por frequência. Entretanto, as médias de aprovação, retenção por rendimento insuficiente e progressão parcial continuam a variar consideravelmente entre os diferentes níveis escolares.

Comparação entre os resultados: Pré-Pandemia e Pandemia

A análise comparativa entre os dados obtidos no período pré-pandemia e durante a pandemia sugere mudanças significativas na composição racial dos alunos matriculados nas escolas. No período pré-pandemia, os dados indicavam uma distribuição mais equilibrada entre alunos brancos, pretos e pardos. Contudo, durante a pandemia, observou-se uma predominância de alunos brancos, o que pode indicar que alunos com melhores condições socioeconômicas, incluindo estrutura familiar, acesso à internet e um ambiente doméstico adequado para os estudos, tiveram maior capacidade de manter a continuidade de suas atividades educacionais.

Essa mudança na composição racial dos alunos parece refletir o impacto das desigualdades socioeconômicas exacerbadas pela pandemia. Presume-se que alunos de famílias mais estruturadas e com melhores recursos econômicos e sociais tiveram mais facilidade em adaptar-se ao ensino remoto, enquanto aqueles de famílias menos favorecidas enfrentaram desafios significativos, dificultando a manutenção de seus estudos.

Além disso, a análise sugere um aumento significativo na taxa de retenção por frequência durante a pandemia. A dificuldade em manter uma frequência regular nas aulas remotas indica contribuição para esse aumento, refletindo os obstáculos enfrentados por muitos alunos na adaptação ao novo formato de ensino. As barreiras tecnológicas e a falta de um ambiente adequado para o estudo parecem ser fatores determinantes para essa realidade. Por outro lado, um aspecto identificado foi que o nível de rendimento de alunos aprovados entre os que conseguiram continuar estudando se manteve instável em relação ao período pré-pandemia, inclusive entre os alunos pretos e pardos. Esse dado sugere que, houve uma eficácia nas adaptações pedagógicas implementadas pelos professores. Esses profissionais mostraram resiliência e competência ao ajustar suas metodologias de ensino às novas condições impostas pela pandemia, garantindo, assim, a qualidade do processo de ensino e aprendizagem.



Por fim, a comparação entre os dados pré-pandemia e durante a pandemia evidencia tanto os desafios quanto as conquistas no campo educacional durante este período. As desigualdades socioeconômicas tiveram um impacto significativo na composição racial dos alunos, na frequência escolar e na estabilidade do rendimento acadêmico. Contudo, a capacidade dos professores em se adaptar às novas condições de ensino remoto demonstrou um esforço notável para mitigar esses desafios. Estas observações são fundamentais para entender o impacto profundo da pandemia na educação e para desenvolver estratégias que possam promover a igualdade e melhorar a capacidade do sistema educacional em futuros contextos de crise.

Discussões

O presente trabalho se destaca ao investigar os efeitos da pandemia no estado de São Paulo, abordando o desempenho das escolas do estaduais no período pré-pandemia e durante a crise de saúde, com base no IDESP e técnicas descritivas, como a clusterização. Silva (2017) também utilizou técnicas de mineração de dados, para prever o desempenho no Ensino médio, destaca-se que o autor utilizou a metodologia CRISP-DM para a realização do processo de mineração de dados.

Já o trabalho de Pinto (2019) realiza uma análise preditiva para identificar os fatores que impactam o Índice de Desenvolvimento da Educação Básica, como escolaridade dos pais e estilo de gestão. Barros (2020) utilizou tanto análise descritiva, como preditiva, em sua pesquisa educacional, entretanto o foco foi a evasão escolar. Primão (2022) também tratou da evasão escolar, mas somente com algoritmos preditivos.

Philot (2021), assim como Barros (2020), utilizou tanto análise descritiva, como preditiva, com alunos do ensino fundamental ao médio, de uma escola privada, para encontrar possíveis indicadores preditivos do desempenho dos alunos, antes que propriamente esses façam as provas. Philot (2021) também fez uso de algoritmo k-means, utilizado nesta pesquisa

Soares (2022), utilizou técnicas preditivas para entender os fatores de influência da qualidade educacional do Maranhão. Assim como Silva (2017) o autor utilizou a metodologia CRISP-DM para a realização do processo de descoberta de conhecimento.

Também utilizaram métodos preditivos os autores: Nunes (2021), para criar materiais didáticos digitais, analisando o rendimento em Língua Inglesa durante a pandemia; Andrelo (2022), visando a analisar de perfis de egressos de escola técnica publica aprovados em exames vestibulares buscando melhorias na gestão educacional e



o desenvolvimento de estratégias educacionais; Queiroz (2022) para avaliar o desempenho da UFRN durante o COVID-19, usando dados de aprovações e reprovações;

Weber Neto et al. (2022) realizaram análises estatísticas sobre impactos da pandemia no ENEM, com resultados sem grandes diferenças, contrastando com as mudanças significativas identificadas nesta pesquisa.

CONSIDERAÇÕES FINAIS

A análise dos clusters, pré-pandemia e pandemia, revela detalhes importantes sobre a distribuição demográfica e o desempenho acadêmico dos alunos. Antes da pandemia, observou-se uma distribuição equilibrada de gênero, com diferenças significativas na distribuição étnica dos alunos, destacando uma predominância de alunos brancos em todos os clusters. Durante a pandemia, houve uma mudança nessa distribuição étnica, com uma concentração maior de alunos brancos e pardos em todos os clusters, indicando um impacto significativo das circunstâncias pandêmicas na demografia estudantil. Quanto aos períodos escolares e desempenho acadêmico, ficou evidente que a maioria dos alunos estava nos Anos Iniciais, seguidos pelos Anos Finais e pelo Ensino Médio. O desempenho acadêmico nos Anos Iniciais foi geralmente mais alto em escolas com um menor número de alunos por unidade escolar, enquanto a retenção por rendimento insuficiente nos Anos Finais foi uma tendência, especialmente em escolas com menos alunos.

Durante a pandemia, os desafios acadêmicos se intensificaram, com uma maior incidência de retenção por frequência e rendimento insuficiente em todos os níveis escolares, especialmente nos Anos Iniciais e Finais. Isso sugere dificuldades de adaptação ao ensino remoto e a necessidade de estratégias educacionais específicas para enfrentar esses desafios. A análise destaca a importância de uma abordagem adaptativa e diferenciada para essa situação, levando em consideração as características demográficas e desafios acadêmicos. Além disso, ressalta a necessidade de colaboração entre escolas, comunidades e autoridades educacionais para implementar estratégias eficazes de apoio aos alunos e garantir uma educação de qualidade e igualdade de oportunidades para todos.

REFERÊNCIAS

ANDRELO, P. F. A. **Mineração de dados educacionais na identificação do perfil dos egressos para apoio à gestão educacional de escola técnica pública**. 2022. 129 f. Dissertação(Programa de Pós-Graduação em Informática e Gestão do Conhecimento) -



Universidade Nove de Julho, São Paulo. Disponível em: <http://biblioteca-tede.uninove.br/handle/tede/3050>. Acesso em: 02 mar. 2024.

BANCO MUNDIAL. **Perdas de aprendizado devido à COVID-19 podem custar a esta geração de estudantes cerca de 17 trilhões de dólares em ganhos ao longo da vida.** Washington, DC: Banco Mundial, 6 dez. 2021. Disponível em: <https://www.worldbank.org/pt/news/press-release/2021/12/06/learning-losses-from-covid-19-could-cost-this-generation-of-students-close-to-17-trillion-in-lifetime-earnings>. Acesso em: 20 out. 2025.

BARROS, T. M. **Um processo orientado a dados para geração de modelo de predição de evasão escolar.** 2020. 116f. Tese (Doutorado em Engenharia Elétrica e de Computação) - Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, 2020. Disponível em: <https://repositorio.ufrn.br/handle/123456789/31933> . Acesso em: 13 fev. 2024.

COLAK, I.; SAGIROGLU, S.; YESILBUDAK, M. Data mining and wind power prediction: a literature review. **Renewable Energy**, v. 46, p. 241–247, 2012. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0960148112001541> . Acesso em: 28 fev . 2024.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, n. 3, p. 37, 15 Mar. 1996.

FERNANDES, Domingos. Para uma teoria da avaliação no domínio das aprendizagens. **Estudos em Avaliação Educacional**, São Paulo, v. 41, n. 19, p. 347-372, 2008.

GIL, Antonio Carlos. **Métodos e Técnicas de Pesquisa Social.** São Paulo: Atlas, 2008.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: concepts and techniques.** Waltham: Elsevier, 2012.

KLEIN, R. Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB). **Ensaio**, [s. l], v. 11, n. 40, p. 283-296, set. 2003. Disponível em: <[file:///C:/Users/Us%C3%BArio/Downloads/38-96-3-PB%20\(1\).pdf](file:///C:/Users/Us%C3%BArio/Downloads/38-96-3-PB%20(1).pdf)> Acesso em: 21 fev. 2024.

NUNES, C. B. de M. P. **A produção de material didático para a formação de professores em análise de dados educacionais.** 2021. 125f. Dissertação (Mestrado Profissional em Inovação em Tecnologias Educacionais) - Instituto Metrópole Digital, Universidade Federal do Rio Grande do Norte, Natal, 2021. Disponível em: <https://repositorio.ufrn.br/handle/123456789/45660>. Acesso em: 28 fev . 2024.

PHILOT, G. M. **Data science aplicado para obtenção de indicadores de desempenho em ambiente escolar remoto.** 2021. Trabalho de Conclusão de Curso (Bacharelado em Física) – Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Engenharia, Guaratinguetá, 2021. Disponível em: <https://repositorio.unesp.br/handle/11449/235329> . Acesso em: 07 mar. 2024.

PINTO, G. da S. **Modelo de análise e predição para identificação dos fatores que influenciam o desempenho escolar na rede de ensino básico: estudo de caso em escolas municipais de Alagoas.** 2019. 84 f. Dissertação (Mestrado em Informática) - Instituto de Computação, Programa de Pós Graduação em Informática, Universidade Federal de Alagoas, Maceió, 2021. Disponível em: <https://www.repositorio.ufal.br/jspui/handle/riufal/7782> . Acesso em: 14 mar. 2024.



PRIMÃO, A. P. **Uso de algoritmos de machine learning para prever a evasão escolar no ensino superior: um estudo no Instituto Federal de Santa Catarina.**

2022. Dissertação (Mestrado Profissional em Administração Universitária) – Universidade Federal de Santa Catarina, Centro Sócio-Econômico, Programa de Pós-Graduação em Administração Universitária, Florianópolis, 2022. Disponível em: : <https://repositorio.ufsc.br/handle/123456789/238320>. Acesso em: 20 mar. 2024.

QUEIROZ, L. J. A. de. **Ferramenta para análise do desempenho acadêmico, utilizando como caso de teste o período sob influência do covid-19.** 2022. 74f.

Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) - Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, 2022. Disponível em: <https://repositorio.ufrn.br/handle/123456789/46107>. Acesso em: 26 mar. 2024.

SÃO PAULO. SECRETARIA DA EDUCAÇÃO DO ESTADO DE SÃO PAULO.

Currículo do Estado de São Paulo: Ciências Humanas e suas tecnologias. São

Paulo: SE, 2011. Disponível em: <https://www.educacao.sp.gov.br/a2sitebox/arquivos/documentos/236.pdf> . Acesso em: 20 out. 2025.

SÃO PAULO. SECRETARIA DA EDUCAÇÃO DO ESTADO DE SÃO PAULO.

IDESP – Índice de Desenvolvimento da Educação do Estado de São Paulo.

Disponível em: <https://dados.educacao.sp.gov.br/story/idesp> . Acesso em: 20 out. 2025.

SÃO PAULO. SECRETARIA DA EDUCAÇÃO DO ESTADO DE SÃO PAULO.

Confira decretos e resoluções de educação implementados durante a pandemia.

São Paulo, 2 jul. 2020. Disponível em: <https://www.educacao.sp.gov.br/confira-decretos-e-resolucoes-de-educacao-implementados-durante-pandemia/>. Acesso em: 20 out. 2025.

SECRETARIA DA EDUCAÇÃO DO ESTADO DE SÃO PAULO. **IDESP – Indicador de Desenvolvimento da Educação do Estado de São Paulo.** Dados Abertos da

Educação, Governo do Estado de São Paulo, 2007. Disponível em: <https://dados.educacao.sp.gov.br/story/idesp>. Acesso em: 20 out. 2025.

SETUBAL, M. A. Os programas de correção de fluxo no contexto das políticas educacionais contemporâneas. **Em Aberto**, Brasília, v. 17, n. 71, p. 9-19, jan. 2000.

SILVA FILHO, R. L. C. **Modelo de análise e predição do desempenho dos alunos dos Institutos Federais de Educação usando o ENEM como indicador de qualidade escolar.** 2017. 137 f. Tese (Doutorado em Ciência da Computação) – Universidade

Federal de Pernambuco, Centro de Informática, Recife, 2017. Disponível em: <https://repositorio.ufpe.br/handle/123456789/28008>. Acesso em: 06 abr. 2024.

SOARES, R. de C. **Mineração de dados para entender os fatores de influência da qualidade educacional do Maranhão.** 2022. 81 f. Dissertação (Programa de Pós-

Graduação em Ciência da Computação/CCET) - Universidade Federal do Maranhão, São Luís, 2022. Disponível em: <https://tede2.ufma.br/jspui/handle/tede/3843>. Acesso em: 01 abr. 2024.

TSAI, H. H. Global data mining: an empirical study of current trends, future forecasts and technology diffusions. **Expert Systems with Applications**, v. 39, n. 9, p. 8172–8181, 2012.

WEBER NETO, N.; SOARES, R. C.; COUTINHO, L. R.; TELES, A. S. A pandemia da COVID-19 impactou o ENEM? Uma análise comparativa de dados dos anos de 2019 e 2020. **RENOTE**, Porto Alegre, v. 20, n. 1, p. 223-232, 2022. Disponível em:

<https://www.seer.ufrgs.br/renote/article/view/126655>. Acesso em: 11 abr. 2024.

