CBIE 2023
CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO
Uma escola para o futuro: Tecnologia e conectividade a serviço da educação

Aplicação de Learning Analytics para Identificação de Tomada de Decisão sobre a Distorção Idade-Série no Brasil

Abílio Nogueira Barros [1], Elyda Laisa Soares Xavier[2], Gabriel Alves[1], Rafael Ferreira Mello[1]

[1]Departamento de Computação – Universidade Federal Rural de Pernambuco (UFRPE)

[2]Universidade Estadual de Pernambuco (UPE)

{abilio.nogueira,rafael.mello,gabriel.alves}@ufrpe.br, elyda.freitas@upe.br

Abstract. *The Age-Grade Distortion Rate (from the Portuguese Taxa de Distorção Idade-Série - TDI) is a key measure for assessing the number of students enrolled in academic years that diverge from their age-appropriate levels. This study leverages Learning Analytics and multi-source data to extract insights to support informed decision-making in this context. Utilizing data from the Basic Education Census, we conducted predictive analytics and feature importance investigation to perform an in-depth analysis. This analysis aims to elucidate the primary factors influencing TDI, enabling a discussion on potential areas for improvement in schools to reduce TDI.*

Resumo. *A Taxa de Distorção Idade-Série (TDI) é um indicador que mede a quantidade de alunos que estejam em um ano curricular diferente de sua idade esperada. Este artigo propõe utilizar Learning Analytics e dados de diferentes fontes para extrair informações relevantes para a tomada de decisão no contexto do TDI. Utilizando fundamentalmente dados do censo da educação básica, foram realizadas predições e análise de importância de características para realizar uma análise detalhada do que mais influência esse indicador e assim prover a discussão sobre o que pode ser mantido ou melhorado em escolas para reduzir o TDI.*

1. Introdução

O sistema educacional brasileiro é permeado por uma série de desafios que impactam diretamente em sua qualidade e paridade entre os entes que compõe o sistema educacional [Schwartzman and Brock 2005, Palomino et al. 2022]. Por exemplo, o Brasil tem um território com proporções continentais e seus mais de 47 milhões de alunos matriculados apenas na educação básica 2022, distribuídos em mais de 178 mil escolas em todo o Brasil[1].

Para auxiliar no acompanhamento de um sistema educacional tão complexo, vários indicadores foram definidos, como taxas de rendimento escolar [Justino 2022], complexidade de gestão escolar[de Andrade et al. 2020], taxa de alfabetização [Bernardi and Luchese 2020], taxa de distorção idade-série (TDI) [NOGUEIRA and Silva 2022], índice de desenvolvimento da edução básica (IDEB) [Rodrigues et al. 2016]. Tais indicadores são calculados e divulgados pelo INEP.

[1]https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar/resultados/

A disponibilização dos dados utilizados para o cálculo desses indicadores permite que sejam realizadas diversas análises, como a desigualdade na educação brasileira [Vitelli et al. 2019], a relação da educação com programas sociais como o Bolsa Família [Santos et al. 2019] e o desempenho educacional dos estudantes através de avaliações como a Prova Brasil [Lacruz et al. 2019].

Neste contexto, a TDI é calculada pelo INEP como uma medida que indica a proporção de alunos em séries diferentes em mais de dois anos da idade correta para a série em questão. O cálculo da TDI envolve a comparação entre a idade dos alunos e a série em que estão matriculados.

Os dados fornecidos pelo INEP podem ser analisados em um contexto histórico através de técnicas de Learning Analytics para extração e processamento desses dados, processo esse utilizado em trabalhos como o de [do Nascimento et al. 2018], que aplicou técnicas de mineração de dados voltados aos dados buscando aplicar modelos de regressão visando encontrar padrões preditivos para evasão escolar. Outro exemplo é [Fonseca and Namen 2016], que de forma mais exploratória visa buscar fornecer apontamentos para a melhoria do sistema educacional baseado no censo da educação básica.

Este estudo propõe utilizar técnicas de Learning Analytics para identificar informações relevantes para a tomada de decisão no contexto de TDI. Para isso, foram utilizados dados do censo da educação básica para identificar as características que podem exercer influência sobre variações no valor desse indicador. Como resultado, é possível identificar pontos de melhoria, que podem ser discutidos gestores educacionais e governamentais, a fim de serem aperfeiçoados.

2. Trabalhos Relacionados

O TDI é um indicador bem consolidado na temática de mapeamento educacional por meio de dados abertos. Para melhor formalizar o cálculo deste indicador, a fórmula a seguir ilustra como a TDI é calculada [Soares and Sátyro 2008]:

$$\text{TDI} = \frac{\text{Número de alunos com idade superior à recomendada para a série}}{\text{Total de alunos matriculados na série}} \times 100$$

Já existem trabalhos que utilizam o monitoramento do TDI para o acompanhamento, não só educacional como o de outras áreas como segurança pública. Como visto em [Ferreira and Teixeira 2018], que retratou como foi possível cruzar os indicadores de violência em localidades que possuíam um alto TDI buscando indicar relação de causa-efeito.

Esse indicador também pode ser usado para avaliar a defasagem entre a educação urbana e rural. Em [NOGUEIRA and Silva 2022] o indicador foi utilizado como ferramenta de acompanhamento de escolas situadas em zona rural e expor assim fatores intra e extraescolares a seus gestores.

Também existem trabalhos que aplicam técnicas de Learning Analytics em microdados educacionais que tratam, desde o agrupamento de estudantes de ensino superior [da Silva Vieira et al. 2022], onde aplicou o algoritmo *k-means* buscando avaliar a qualidade do ensino superior brasileiro. Como em [Viana et al. 2022], que propôs uma forma de classificação de evasão e graduação aplicados a cursos de computação da Universidade Federal do Piauí (UFPI).

Esses trabalhos serviram para nortear a forma com que esse indicador já foi utilizado e para fortalecer o entendimento do processo já utilizado por outros grupos de pesquisa. Não encontramos trabalhos específicos sobre a utilização de Learning Analytics para tomada de decisão sobre o TDI. Com isso, nosso trabalho visa apresentar um panorama de todas as escolas brasileiras entre 2018 a 2022, pois anos esse que foram anos e transições tanto governamentais quanto sociais, com a educação remota de emergência durante a pandemia [de Souza et al. 2021].

3. Metodologia

O presente estudo emprega técnicas de Learning Analytics (LA) na base de dados do censo educacional para alcançar o objetivo. Inicialmente, foi construída a base de dados, compreendendo todas as variáveis necessárias para a investigação. Através da aplicação de LA visando avaliar diversas combinações de características visando aprimorar a precisão preditiva do modelo.Ao final dos experimentos foram realizadas análise dos resultados, destacando a identificação das características mais relevantes para aquele resultado. Cada etapa está descrita em mais detalhes nas próximas seções.

3.1. Aquisição e Processamento dos Dados

Os dados utilizados para esse projeto foram todos adquiridos do INEP respeitando a faixa temporal definida por esse estudo. As bases necessárias foram o TDI ao nível de entidade escolar [2] e o censo escolar da educação básica [3].

Para o processamento do censo escolar foi realizada a extração e transformação de dados, com base no que foi descrito em [Barros et al. 2022, de Albuquerque et al. 2022], a fim de unificar os anos e assim preparar uma base para os algoritmos de LA. O processo necessário para a criação da base envolveu a remoção de colunas que tivessem uma quantidade de valores nulos que ultrapassem 80% dos dados [Barros et al. 2022, de Albuquerque et al. 2022]. Além disso, foram removidas as colunas de identificação direta da instituição escolar, como colunas que falassem sobre localidade geográfica [Barros et al. 2023]. Ficando assim a base final com cerca de 335 atributos, a lista de colunas removidas pode ser encontrada nesse endereço [4]. Foi necessário criar apenas uma coluna não original do censo para que pudesse se adequar a próxima etapa do estudo, sendo ela DIAS_LETIVOS, pois se trata da diferença entre as datas informadas no censo para o cômputo em dias, criando assim uma coluna quantitativa onde anteriormente haviam duas do tipo data.

Para a padronização dos arquivos originais do indicador TDI, coluna alvo para esse estudo, o processo a extração dos arquivos focou no ensino fundamental anos iniciais, por questões de limitação de espaço no artigo. Com isso, apenas escolas que apresentaram o resultado desse indicador em seu respectivo ano. A partir disto, uniu-se os resultados do censo em que existia correspondência direta entre instituição e ano, formando assim a base entre o resultado que visamos entender com as informações escolares dispostas no censo. Formando assim uma base de pouco mais de 533 mil registros.

[2]https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/indicadores-educacionais/taxas-de-distorcao-idade-serie

[3]https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/censo-escolar

[4]https://www.zenodo.org/record/6666613

Assim como outros indicadores já citados nesse trabalho, o TDI é expresso como o percentual de alunos com matrículas escolares que estão atrasados em relação à idade escolar adequada.Neste caso, quanto menor sua porcentagem, melhor é a situação da instituição. Em outras palavras, escolas com TDI próximo de 0% tem menos alunos com a distorção de idade e série.

3.2. Análise dos Dados

Após a preparação da base de dados, foi definido que seria utilizado algoritmos de regressão, visto que o valor alvo é numérico, para predizer o TDI. Para essa etapa do projeto foi utilizado um algoritmo caixa branca a fim de aumentar a explicabilidade de seus resultados, mais especificamente utilizamos o *Decision Tree Regressor*[Mahbooba et al. 2021]. Para buscar a melhoria no algoritmo escolhido, foi utilizada a otimização de hiperparâmetros, visando melhorar o algoritmo antes de sua execução com base nos dados que vão ser executados, podendo assim gerar resultados mais satisfatórios[Luo 2016]. A fim de gerar o melhor modelo possível, utilizamos a biblioteca *GridSearchCV*, que implementa o algoritmo Grid Search com o objetivo de otimizar os parâmetros do modelo [Huang et al. 2012].

Com a escolha dos melhores parâmetros para aplicação do modelo, foram realizados três configurações de experimentos nos dados coletados:

Experimento 1: Nessa etapa, todas as colunas da base de dados foram mantidas.
Experimento 2: Foram selecionadas apenas as características que informavam sobre matrícula, sendo assim todas de teor quantitativo.
Experimento 3: Utilização das características qualitativas do censo, excluindo as matrículas.

Para cada experimento, os classificadores foram executados com a configuração determinada pelo otimizador de hiperparâmetros. Após a execução foram escolhidas as métricas tradicionais para avaliação de algoritmos de regressão para medir o desempenho dos modelos desenvolvidos:

- Correlação de Pearson: Mede a força e direção da correlação linear entre duas variáveis contínuas, variando de -1 a 1[Stoean et al. 2013];
- Correlação de Spearman: Avalia a relação monotônica entre duas variáveis, atribuindo classificações ordinais aos dados[Budach et al. 2022];
- RMSE (Root Mean Square Error): Métrica de erro que calcula a diferença média entre os valores previstos e os valores reais, onde quando menor o valor, melhor o desempenho do modelo[Arghandabi and Shams 2020].

Além disto, no contexto deste estudo, foram selecionadas as dez melhores características apontadas por cada um dos algoritmos em seus respectivos experimentos. Para o levantamento dessas melhores características foi utilizado o método *feature_importances_*[5], as características são calculadas com base na média e desvio padrão das do acúmulo da diminuição das impurezas geradas nas classes das árvores.

O objetivo dessa etapa não é de fato selecionar apenas as dez colunas para novas rodadas de processamento em busca de prever o TDI de escolas em novos anos, e sim

[5]https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html

poder investigar as características apontadas pelo modelo com maior potencial de caracterizar o TDI.

Junto à descrição do que cada característica representa, teremos uma coluna informando os valores das escolas que apontaram uma melhor situação, dado que o TDI menor indica uma melhor situação educacional, e as que estão no grupo em pior situação, instituições que estejam superior à média de TDI. Para a análise a seguir, o TDI médio da educação básica foi de 11.71%, dividindo assim a base de dados de mais de 500 mil registros entre TDI inferior à média (indicando um bom índice) e TDI superior à média (indicando um índice ruim).

Em resumo, para a avaliação dos resultados, foi aplicada a seguinte abordagem:

1. Calcular a média do TDI das entidades escolas;
2. Após definido o valor médio, a base de dados é adicionada a coluna se a entidade escolar está abaixo (valor de TDI menor que a média, situação melhor) ou acima (valor de TDI maior que a média, situação pior);
3. São selecionadas as dez características mais importantes de cada experimento, as colunas podem ser de dois tipos:
    - Quantitativos: Onde será exibida a quantidade média e o desvio padrão da característica.
    - Categórica: Onde será exibida a quantidade de escolas em cada categoria daquela característica.

4. Resultados

Como cada experimento possuía diferente conjunto de dados, foram gerados os resultados com base em cada rodada de execução do algoritmo e seus respectivos dados. As tabelas 1, 2 e 3 mostram os resultados para a melhor execução do modelo após a utilização do grid search nos parâmetros[6].

Tabela 1. Resultados das Métricas — Experimentos 1, 2 e 3

		Métrica		
Experimento	Descrição	Pearson	Spearman	RMSE
1	Base completa	0,786225	0,740435	8,571647
2	Base matrículas	0,797232	0,735176	8,358507
3	Base atributos	0,582022	0,510149	11,289482

Após a execução dos algoritmos é calculada a importância das características. Essa abordagem permite identificar quais características têm maior contribuição para a previsão e, potencialmente, descartar características menos relevantes, simplificando o modelo e melhorando sua capacidade de generalização. Em cada experimento as colunas categóricas são destacadas a fim de exemplificar como aquele resultado foi aplicado na tabela relacionado a cada experimento, buscando fornecer uma maior interpretabilidade da característica que foi destacada.

[6]Não foi possível mostrar o resultado de todos os parâmetros devido à limitação de espaço do artigo

Ao observarmos os resultados na Tabela 2, conseguimos observar que do grupo de entidades escolares que possuem um TDI inferior à média, temos mais matriculas, com quantidades maiores de turmas, em todas as etapas do ensino fundamental. A tabela também aponta que as entidades escolares que possuem internet disponível na entidade escolar são 80% dentre os posicionados abaixo da média. Já as entidades de dependência pública representam 90% das escolas que possuem esse indicador mais elevado que a média. Observa-se também que, embora esteja consideravelmente distante do público-alvo deste estágio educacional, há uma significativa quantidade de matrículas de estudantes com 18 anos ou mais. Essas matrículas indicam que esses indivíduos encontram-se além da faixa etária prevista para a conclusão do ciclo escolar completo. Já nesse primeiro experimento podemos destacar também que a quantidade de escolas públicas é mais de 90% das instituições com valores de TDI acima da média nessa faixa, necessitando assim um maior olhar ao sistema público de ensino.

Tabela 2. Importância das Características — Experimento 1

			TDI	
Característica	Descrição	Importância	Inferior	Superior
QT_MAT_BAS_6_10	Qtd de matrículas 6 a 10 anos	0.24	148.25 (160.94)	93.53 (120.36)
QT_MAT_BAS_11_14	Qtd de matrículas 11 a 14 anos	0.19	65.97 (112.12)	57.99 (101.51)
IN_INTERNET	Possuem acesso a internet	0.016	82.26 (17.74)	56.62 (43.38)
QT_MAT_BAS_15_17	Qtd de matrículas 15 a 17 anos	0.09	22.34 (65.55)	18.50 (47.75)
QT_MAT_FUND_AI	Qtd de matrículas no fundamental anos iniciais	0.05	153.74 (168.38)	110.46 (139.38)
QT_MAT_FUND_AF	Qtd de matrículas no fundamental anos finais	0.04	65.44 (118.97)	53.23 (113.78)
QT_MAT_BAS_BRANCA	Qtd de matrículas de alunos declarados brancos	0.04	108.03 (163.59)	36.78 (81.35)
TP_DEPENDENCIA	Tipo de dependência da instituição	0.03	71.41 (28.58)	92.87 (7.13)
IN_ENERGIA	Utiliza fornecimento de energia pública	0.02	97.80 (2.20)	88.08 (11.92)
QT_MAT_BAS_18_MAIS	Qtd de matrículas 18 anos ou mais.	0.02	9.85 (40.14)	16.39 (45.99)

Avaliando agora os resultados do experimento 2, apresentados na tabela 3. Podemos observar que em números, que apenas 3 das 10 características mais importantes apontadas pelo modelo indicam que mais matrículas trazem um TDI menor que a média geral dos casos analisados. Trazendo assim a leitura que o problema não está diretamente ligado a quantidade de matrículas da instituição nem a quantidade de diferentes faixas etárias de matrículas na mesma instituição. Ainda na mesma tabela vemos que a questão da declaração de etnia se destacando, onde instituições com uma maior quantidade de alunos declarados brancos estariam abaixo da média do indicador, obtendo assim um melhor resultado, indicativo esse que pode remeter a questões mais uma vez sobre o sistema escolar público/privado.

Tabela 3. Importância de Característica - Experimento 2

			TDI	
Característica	Descrição	Importância	Inferior	Superior
QT_MAT_BAS_6_10	Qtd de matrículas 6 a 10 anos	0.28	148.25 (160.94)	93.53 (120.36)
QT_MAT_BAS_11_14	Qtd de matrículas 11 a 14 anos	0.22	65.97 (112.12)	57.99 (101.51)
QT_MAT_BAS_BRANCA	Qtd de matrículas de alunos declarados brancos	0.15	108.03 (163.59)	36.78 (81.35)
QT_MAT_FUND_AI	Qtd de matrículas no fundamental anos iniciais	0.09	153.74 (168.38)	110.46 (139.38)
QT_MAT_BAS_15_17	Qtd de matrículas 15 a 17 anos	0.07	22.34 (65.55)	18.50 (47.75)
QT_MAT_FUND_AF	Qtd de matrículas no fundamental anos finais	0.04	65.44 (118.97)	53.23 (113.78)
QT_MAT_EJA_FUND	Qtd de matrículas na modalidade EJA no ensino fundamental.	0.03	6.61 (30.31)	14.98 (46.0)
QT_MAT_BAS_18_MAIS	Qtd de matrículas 18 anos ou mais.	0.03	9.85 (40.14)	16.39 (45.99)
QT_MAT_MED	Qtd de matrículas no ensino médio.	0.01	18.06 (69.54)	5.45 (41.91)
QT_MAT_BAS_PRETA	Qtd de matrículas de alunos declarados pretos.	0.01	7.43 (14.98)	8.36 (19.27)

Por fim, temos os resultados expressos na Tabela 4 e nesse resultado podemos destacar que a maioria de características diretamente ligadas a estrutura escolar sejam mais relevantes. Foi destacado o uso de internet tanto no geral como diretamente ligada ao processo de ensino e aprendizagem, demonstrando que a utilização dessa tecnologia, que também requer outras atreladas a seu uso, dando assim um destaque a temática de tecnologias na educação que deve ser considerado por gestores e professores. A quantidade de dias letivos também e mostrada, onde uma maior quantidade de dias traria um TDI abaixo da média. A quantidade turmas e de docentes em uma maior quantidade pode acarretar valores de indicadores mais baixos, levantando a discussão sobre quantidade de alunos por turma e um maior quadro de docentes.

5. Implicações Práticas

As implicações práticas destes resultados podem refletir em diferentes contextos.A primeira, e principal implicação, está relacionada à tomada de decisão para políticas públicas. Por exemplo, a identificação que a quantidade de matrículas por escola influencia a TDI pode levar a decisões de redistribuição de alunos. Ainda neste contexto, o experimento 3 lista uma grande quantidade de características que estão impactando negativamente o TDI que podem ser direcionadoras de políticas públicas eficientes.

O segundo ponto relevante das análises estão relacionadas ao contexto dos resultados. As tabelas apontam para algumas características determinantes para maiores (ou

Tabela 4. Feature importance experimento 3

			TDI	
Característica	Descrição	Importância	Inferior	Superior
IN_INTERNET	Possuem acesso a internet	0.22	82.26 (17.73)	56.62 (43.38)
IN_ENERGIA_REDE_PUBLICA	Utiliza fornecimento de energia pública	0.10	97.80 (2.20)	88.08 (11.92)
TP_DEPENDENCIA	Tipo de depência da instituição	0.06	71.41 (28.59)	92.87(7.13)
IN_LOCAL_FUNC_SOCIOEDUCATIVO	A escola disponibiliza atendimento socioeducativo	0.04	0.03 (99.97)	0.23 (99.77)
QT_TUR_FUND_AI	Número de Turmas de Ensino Fundamental - Anos Iniciais	0.03	7.08 (6.39)	4.82 (5.62)
TP_INDIGENA_LINGUA	Língua indígena que é ministrada.	0.03	99.04 (0.96)	95.56 (4.42)
DIAS_LETIVOS	Quantidade de dias letivos.	0.05	312.08 (19.66)	305.49 (25.27)
QT_DOC_FUND_AI	Número de Docentes do Ensino Fundamental - Anos Iniciais	0.03	8.98 (8.37)	5.45 (6.52)
IN_INTERNET_APRENDIZAGEM	Possuem acesso à internet para uso nos processos de ensino e aprendizagem	0.02	45.50 (54.50)	19.73 (80.27)
CO_LINGUA_INDIGENA_1	Escolha da língua indígena	0.03	99.04 (0.96)	95.56 (4.44)

seja, piores) TDIs para escolas como a restrição no acesso à internet e energia (por exemplo). Além disto, grupos desfavorecidos também influenciam bastante no resultado. Por fim, os resultados apresentados também podem ser utilizados por gestores escolares podem validar aspectos administrativos da escola, como quantidade de alunos por turma ou utilização de laboratórios.

6. Conclusão

Os resultados apresentados neste trabalho mostram que o *Decision Tree Regressor* consegue identificar as variáveis que mais impactaram o TDI. A divisão das bases de dados, para assim gerar diferentes experimentos, pôde trazer uma melhor visão dos resultados, visto que a quantidade de matrículas foi um dos elementos mais apontados desde a primeira rodada de execução do algoritmo. Levantando assim que escolas com uma maior quantidade de turmas e matrículas tendem a possuir um TDI menos elevado em relação as com menos matrículas e turmas, não só apenas da faixa escolar levantado no estudo.

Com vistas à continuidade desta pesquisa, almejamos ampliar a análise para abrangeros anos finais do ensino fundamental e o ensino médio.Adicionalmente, planejamos efetuar uma comparação das características mais significativas em cada etapa educacional, acompanhando seus resultados por entidade escolas. Paralelamente, aspiramos validar tais resultados junto a gestores educacionais, que, devido ao seu conhecimento especializado na área, poderão enriquecer a avaliação dos resultados apontados por este algoritmo ou futuras otimizações do mesmo [Spanol et al. 2022].

Referências

Arghandabi, H. and Shams, P. (2020). A comparative study of machine learning algorithms for the prediction of heart disease. *Int J Res Appl Sci Eng Technol. https://doi. org/10.22214/ijraset*.

Barros, A. N., Alencar, A., Nascimento, A., de Albuquerque, A. F., and Mello, R. F. (2022). Elaboração do conjunto de dados agregados do censo da educação básica. In *Anais do IV Dataset Showcase Workshop*, pages 35–45. SBC.

Barros, A. N., de Albuquerque, A. F., Alencar, A., Mello, R. F., Alves, G., and Bittencourt, I. M. (2023). Arquitetura de dados educacionais como plataforma para governo inteligente-utilizando dados abertos para apoio à gestão educacional baseada em evidências. In *Anais do XI Workshop de Computação Aplicada em Governo Eletrônico*, pages 130–140. SBC.

Bernardi, M. C. and Luchese, T. Â. (2020). A taxa de alfabetização de antônio prado, rio grande do sul (1895-1920). *Revista Educação em Questão*, 58(56).

Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., and Harmouch, H. (2022). The effects of data quality on machine learning performance. *arXiv preprint arXiv:2207.14529*.

da Silva Vieira, A., Bertolini, D., and Schwerz, A. L. (2022). Análise do desempenho no enade dos concluintes de computaçao usando técnica de agrupamento. In *Anais do XXXIII Simpósio Brasileiro de Informática na Educação*, pages 834–845. SBC.

de Albuquerque, A. F., Barros, A. N., Alencar, A., Nascimento, A., Bittencourt, I. M., and Mello, R. F. (2022). Dataset de estimativas populacionais desagregada por município e idade 2014-2020. In *Anais do IV Dataset Showcase Workshop*, pages 25–34. SBC.

de Andrade, M. C. B., Silva, L. F., Fecury, A. A., de Oliveira, E., Dendasck, C. V., de Araújo, M. H. M., da Souza, K. O., da Silva, I. R., de Medeiros Moreira, E. C., Pascoal, R. M., et al. (2020). Indicadores de complexidade de gestão em escolas públicas e privadas de duas cidades do estado do amapá entre 2014 e 2018. *Research, Society and Development*, 9(9):e856998112–e856998112.

de Souza, G. H. S., Jardim, W. S., Marques, Y. B., Junior, G. L., dos Santos, A. P. S., and de Paula Liberato, L. (2021). Educação remota emergencial (ere): um estudo empírico sobre capacidades educacionais e expectativas docentes durante a pandemia da covid-19. *Research, Society and Development*, 10(1):e37510111904–e37510111904.

do Nascimento, R. L. S., da Cruz Junior, G. G., and de Araújo Fagundes, R. A. (2018). Mineração de dados educacionais: Um estudo sobre indicadores da educação em bases de dados do inep. *RENOTE*, 16(1).

Ferreira, V. B. and Teixeira, E. C. (2018). O impacto da distorção idade-série sobre a criminalidade nos municípios de minas gerais. *Revista Brasileira de Segurança Pública*, 12(2):269–291.

Fonseca, S. O. d. and Namen, A. A. (2016). Mineração em bases de dados do inep: uma análise exploratória para nortear melhorias no sistema educacional brasileiro. *Educação em Revista*, 32:133–157.

Huang, Q., Mao, J., and Liu, Y. (2012). An improved grid search algorithm of svr parameters optimization. In *2012 IEEE 14th International Conference on Communication Technology*, pages 1022–1026. IEEE.

Justino, M. R. (2022). A relação do esforço docente e da infraestrutura escolar nas taxas de rendimento escolar: uma análise para a cidade do natal no ano de 2019. B.S. thesis, Universidade Federal do Rio Grande do Norte.

Lacruz, A. J., Américo, B. L., and Carniel, F. (2019). Indicadores de qualidade na educação: análise discriminante dos desempenhos na prova brasil. *Revista brasileira de educação*, 24:e240002.

Luo, G. (2016). A review of automatic selection methods for machine learning algorithms and hyper-parameter values. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 5:1–16.

Mahbooba, B., Timilsina, M., Sahal, R., and Serrano, M. (2021). Explainable artificial intelligence (xai) to enhance trust management in intrusion detection systems using decision tree model. *Complexity*, 2021:1–11.

NOGUEIRA, M. D. O. E. and Silva, L. C. (2022). Escolarização em áreas rurais: a distorção idade-série na ótica dos gestores. *Estudos em Avaliação Educacional*, 33.

Palomino, P., Falcao, T. P., Medeiros, R., Uehara, M., Bittencourt, I., and Mello, R. F. (2022). Plataformas de dados educacionais: Análise com foco no plano nacional de educaçao. In *Anais do I Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil*, pages 60–68. SBC.

Rodrigues, E. C. et al. (2016). Indicadores educacionais e contexto escolar: uma análise das metas do ideb. *Estudos em Avaliação Educacional*, 27(66):662–688.

Santos, M. C. S., Delatorre, L. R., Ceccato, M. d. G. B., and Bonolo, P. d. F. (2019). Programa bolsa família e indicadores educacionais em crianças, adolescentes e escolas no brasil: revisão sistemática. *Ciência & Saúde Coletiva*, 24:2233–2247.

Schwartzman, S. and Brock, C. (2005). Os desafios da educação no brasil. *Rio de Janeiro: Nova Fronteira*, 1320.

Soares, S. and Sátyro, N. (2008). O impacto de infra-estrutura escolar na taxa de distorção idade-série das escolas brasileiras de ensino fundamental: 1998 a 2005. Technical report, Texto para Discussão.

Spanol, M., Oliveira, E., Alves, G., Bittencourt, I. M., Falcao, T. P., and Mello, R. F. (2022). Uso de agrupamento para avaliação de desempenho educacional e apoio à gestão em áreas de investimento. In *Anais do XXXIII Simpósio Brasileiro de Informática na Educação*, pages 944–955. SBC.

Stoean, C., Preuss, M., and Stoean, R. (2013). Ea-based parameter tuning of multimodal optimization performance by means of different surrogate models. In *Proceedings of the 15th annual conference companion on Genetic and evolutionary computation*, pages 1063–1070.

Viana, F. S., Santana, A. M., and Rabêlo, R. d. A. L. (2022). Avaliaçao de classificadores para prediçao de evasao no ensino superior utilizando janela semestral. In *Anais do XXXIII Simpósio Brasileiro de Informática na Educação*, pages 908–919. SBC.

Vitelli, R. F., Fritsch, R., and da Silva, R. D. (2019). A desigualdade brasileira revelada pelo resultado de indicadores educacionais.