

Árvore de Classificação Para Estimação de Perfis de Estudantes que Participaram da Prova SAEB de Língua Portuguesa no Ano de 2019

Pedro dos S. Costa Junior[†], Carla Dominique S. Vasconcelos¹, Ana Vitória C. da Silva¹, Neylan L. Dias¹, Simone de A. D. Leal¹, Edcarlos V. da Silva¹

¹ Universidade Federal do Amapá (UNIFAP); Macapá – Amapá, Brasil.

Resumo: O SAEB é uma avaliação nacional em larga escala do Brasil que mede a qualidade do ensino da Educação Básica do Brasil, abrangendo Língua Portuguesa e Matemática. O objetivo deste estudo foi o de estimar perfis de estudantes amapaenses que estão abaixo da média nacional em Língua Portuguesa, que participaram da prova SAEB de 2019. A pesquisa utilizou a metodologia de árvore de decisão para estimar a probabilidade de o aluno estar abaixo da média nacional, empregou-se o modelo CHAID para analisar o cruzamento das notas com variáveis sociodemográficas. Observou-se que 64,7% dos estudantes do Amapá estavam com notas abaixo da média nacional (278,4) em Língua Portuguesa. A árvore de classificação revelou que alunos de escolas públicas de áreas rurais têm uma chance de 83,2% de terem notas abaixo da média em Língua Portuguesa, sendo esse o perfil de maior probabilidade. Como conclusão verificou-se que esse estudo destacou os perfis com maior probabilidade de baixo rendimento em Língua Portuguesa, o conhecimento desses perfis contribui para elaboração de políticas públicas com foco nos alunos em maiores vulnerabilidades.

Palavras-chave: SAEB; Árvore de decisão; Amapá; Educação; Avaliação.

Classification Tree for Estimating Student Profiles Who Participated in the SAEB Portuguese Language Exam in 2019

Abstract: The SAEB is a large-scale national assessment in Brazil that measures the quality of education in Brazilian Basic Education, covering Portuguese Language and Mathematics. The objective of this study was to estimate profiles of students from Amapá who are below the national average in Portuguese Language and who participated in the 2019 SAEB exam. The research used decision tree methodology to estimate the probability of a student being below the national average, employing the CHAID model to analyze the intersection of scores with sociodemographic variables. It was observed that 64.7% of Amapá students had scores below the national average (278.4) in Portuguese Language. The classification tree revealed that students from public schools in rural areas have an 83.2% chance of having scores below the national average in Portuguese Language, representing the profile with the highest probability. In conclusion, this study highlighted the profiles with the highest likelihood of low performance in Portuguese Language, and understanding these profiles contributes to the development of public policies focused on students in greater vulnerability.

Keywords: SAEB; Decision tree; Amapá; Education; Assessment.

[†] Autor correspondente: pedrodoss028@gmail.com

Manuscrito recebido em: 07/06/2024
Manuscrito revisado em: 24/10/2024
Manuscrito aceito em: 25/10/2024

Introdução

O Sistema de Avaliação da Educação Básica (SAEB) é uma das principais ferramentas de avaliação educacional no Brasil. Gerenciado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), o SAEB tem como objetivo avaliar a qualidade do ensino nas escolas públicas e privadas do país. A avaliação é aplicada regularmente aos alunos do 5º e 9º anos do Ensino Fundamental, bem como aos alunos do 3º e 4º anos do Ensino Médio. O SAEB consiste em testes padronizados que avaliam o desempenho dos alunos em duas áreas principais: língua portuguesa e matemática. Os questionários são desenvolvidos e aplicados pelo INEP, uma autarquia vinculada ao Ministério da Educação (MEC). Essas avaliações são elaboradas de acordo com as diretrizes curriculares nacionais e têm como finalidade medir o domínio dos estudantes em relação aos conteúdos e habilidades esperadas para cada série.

Por meio dos instrumentos utilizados, provas e questionários, sistematicamente pelo INEP, o SAEB permite que as esferas municipais, estaduais e federais de governo avaliem a qualidade da educação praticada em sua jurisdição com a finalidade de oferecer subsídios para a construção, o acompanhamento e o aperfeiçoamento de políticas com amparo nessas evidências. Ainda, as médias de desempenho obtidas no SAEB, juntamente com os dados sobre aproveitamento obtidos no Censo Escolar, compõem o IDEB (Coelho; Lopes; Costa, 2022).

Essas avaliações são realizadas bienalmente e têm como objetivo principal diagnosticar a situação da educação básica nacional, incluindo fatores influenciadores do desempenho dos alunos. Os resultados colaboram para a formulação e/ou adaptação de políticas públicas em todas as esferas governamentais, com intuito de proporcionar uma melhor qualidade de ensino para os cidadãos. Para esse estudo foram utilizados os resultados da avaliação SAEB, pois tem como um dos focos a análise do desempenho dos alunos egressos do Ensino Médio (Silva; Silva; Martins, 2015).

Os resultados do SAEB têm múltiplas utilidades. Eles fornecem uma base para a criação de políticas públicas educacionais, destacando áreas que demandam mais atenção e investimentos. Além disso, tanto os resultados individuais quanto os institucionais podem ser usados por escolas e gestores educacionais para identificar pontos fortes e fracos, direcionando ações para aprimoramento pedagógico. Uma das análises possíveis com os dados do SAEB é o estudo do desempenho dos alunos, relacionando os resultados das provas de Matemática ou Língua Portuguesa com variáveis sociodemográficas. Isso permite identificar quais perfis de alunos obtêm as menores e as maiores notas na avaliação, ajudando a detectar disparidades e revelando os grupos mais vulneráveis ao insucesso escolar.

Para realizar essa análise e identificar esses perfis, foi obtido do site do INEP o banco de dados (microdados) referente ao ano de 2019. A partir desses dados, foram selecionadas as informações dos estudantes do Amapá, incluindo as notas da prova de Língua Portuguesa e variáveis sociodemográficas para análise. Foram realizadas estatísticas descritivas e utilizado um algoritmo de árvore de decisão para calcular as probabilidades de os estudantes apresentarem perfis com notas abaixo da média nacional no ano de 2019.

Nesse sentido, este trabalho tem como objetivo analisar o desempenho escolar dos alunos do Ensino Médio do Amapá na disciplina de Língua Portuguesa, com foco na identificação dos perfis de estudantes com maior probabilidade de apresentar notas abaixo da média nacional do SAEB referente ao ano de 2019.

Materiais e Métodos

Os dados foram obtidos diretamente do site do INEP, e corresponderam ao banco de dados do SAEB, disponibilizados gratuitamente pela plataforma para que pesquisadores possam ter acesso e desenvolver estudos. Este estudo consistiu em uma pesquisa de abordagem quantitativa (Silva; Menezes, 2005) com uso de banco de dados secundários (Lapei, 2024).

O Brasil, em 2019, apresentou nota média na prova de Língua Portuguesa de 278,4 pontos (INEP, 2024). Essa foi a referência para que o estudante do Amapá, do Ensino Médio, fosse classificado em relação a sua nota pessoal e a média nacional. Desta forma, procedeu-se com a seguinte classificação:

- Nota de Língua Portuguesa abaixo da média nacional de 278,4 pontos: Aluno Abaixo da Média Nacional;
- Nota de Língua Portuguesa acima da média nacional de 278,4 pontos: Aluno Acima da Média Nacional.

Optou-se pela modalidade de dados conhecida como microdados, que reúnem um conjunto de informações detalhadas relacionadas às pesquisas, aos exames e avaliações do INEP, filtrado para as provas de Língua Portuguesa do SAEB, com foco nos estudantes do Amapá. Esses microdados são informações individuais dos alunos do Ensino Médio que fizeram a prova no ano de 2019 (INEP, 2024). Esses dados foram cruzados com variáveis sociodemográficas dos alunos como: zona de residência (capital e interior), área de residência (urbana e rural), sexo (masculino e feminino), acesso a rede Wi-fi (sim e não), raça (branco, pardo/preto e indígena).

O cruzamento de dados foi feito usando o algoritmo Árvore de Decisão, modelo *Chi-Square Automatic Interaction Detector* – CHAID, que analisa interações de uma variável resposta (Notas dos alunos em Língua Portuguesa) com prováveis variáveis explicativas (sexo, raça, local de residência, etc) na forma de probabilidades (Pestana; Gageiro, 2009). O modelo identifica um perfil com base em todas as variáveis incluídas e estabelece uma probabilidade de ocorrência daquele perfil com base na variável resposta. O método CHAID foi utilizado para analisar quais perfis de alunos do Ensino Médio do Amapá têm maior probabilidade de estar abaixo da média nacional na prova de Língua Portuguesa do SAEB, identificando assim os perfis mais vulneráveis para a prova. O algoritmo foi implementado utilizando o pacote estatístico SPSS versão 26.

A população em estudo consistiu no censo total de estudantes do Amapá do Ensino Médio que participaram da prova de Língua Portuguesa em 2019, amostra incluiu 5.876 estudantes, sendo 60% reservado para treinamento do algoritmo e 40% para o teste efetivo do algoritmo CHAID para estimar as árvores de decisão com os perfis de probabilidades, as árvores foram apresentadas conforme a saída gráfica do modelo matemático, na forma de caixas e nós (Pestana; Gageiro, 2009).

Como o estudo utilizou dados secundários disponibilizados publicamente na internet, não foi necessário registrar a pesquisa em um Comitê de Ética em Pesquisa. Isso se deve ao fato de que, embora os dados secundários contenham informações individuais, não há identificação pessoal dos alunos. Além disso, a pesquisa não envolveu interação direta com pessoas, tornando desnecessário o registro do projeto de pesquisa em um órgão de fiscalização.

Resultados e Discussões

A primeira árvore de decisão apresenta 9 nós sendo 6 deles considerados nós terminais (3, 4, 5, 6, 7 e 8) a amostra incluída na árvore teve tamanho $n = 2369$ e analisou a proficiência dos estudantes (abaixo da média e acima da média) com significância para as seguintes variáveis: raça/cor (pardo/preta/amarela/indígena e branco), ao concluir o ensino médio (somente continuar estudando, continuar estudando/trabalhando, somente trabalhar ou ainda não sei).

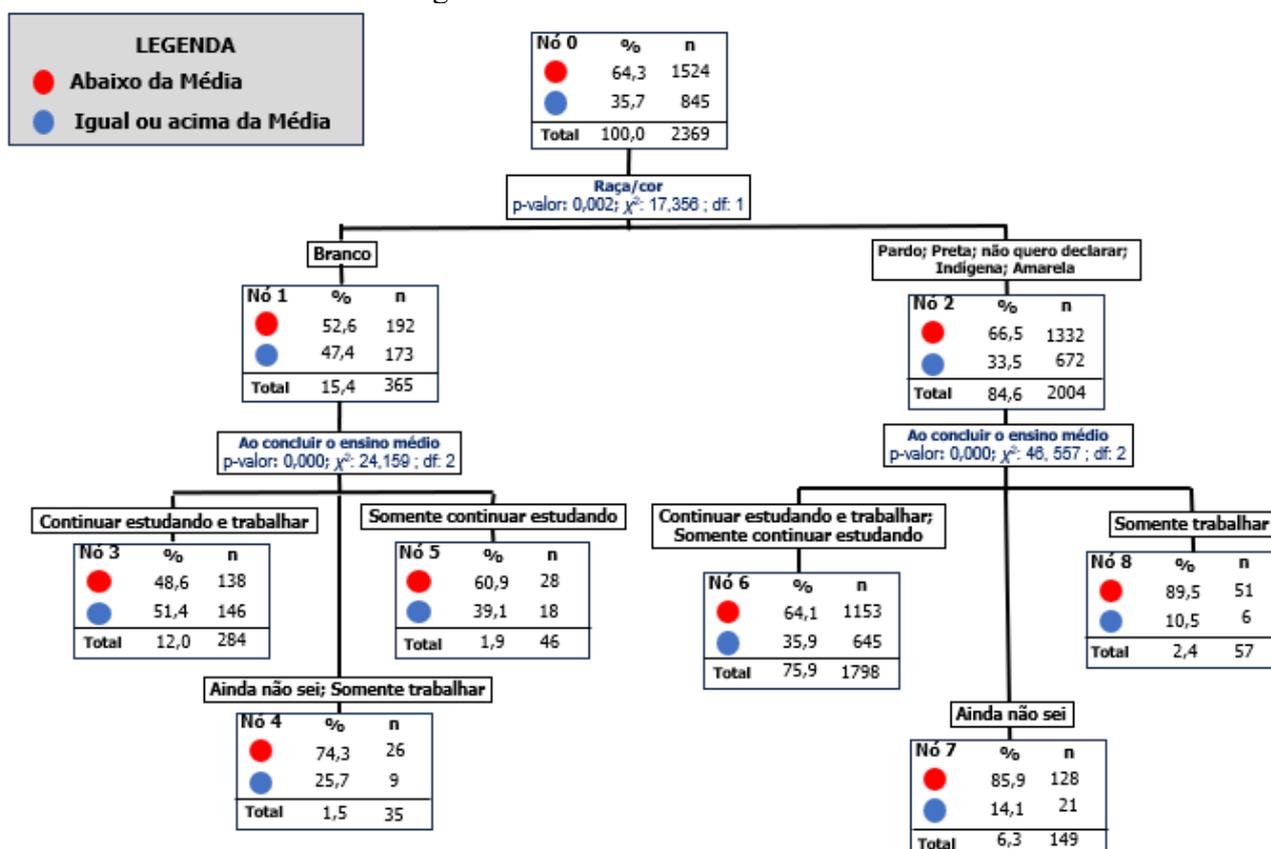
A Figura 1 apresenta os perfis de probabilidades para identificar estudantes do Amapá com mais chances de estarem abaixo da média nacional em Língua Portuguesa na prova do SAEB.

Os perfis identificados foram:

- a) Do nó 0 ao nó 8: estudantes de raça/cor pardo/preta/amarela/indígena ao concluir o ensino médio, somente trabalhar, possuem 89,5% de probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB;
- b) Do nó 0 ao nó 7: estudantes de raça/cor pardo/preta/amarela/indígena ao concluir o ensino médio, ainda não sabe, possuem 85,9% de probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB;
- c) Do nó 0 ao nó 4: estudantes de raça/cor branco ao concluir o ensino médio, somente trabalhar, possuem 74,3% de probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB;
- d) Do nó 0 ao nó 6: estudantes de raça/cor pardo/preta/amarela/indígena ao concluir o ensino médio, continuar estudando e trabalhando e somente continuar estudando, possuem 64,1% de probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB.
- e) Do nó 0 ao nó 5: estudantes de raça/cor branco ao concluir o ensino médio, somente continuar estudando, possuem 60,9% de probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB.
- f) Do nó 0 ao nó 3: estudantes de raça/cor branco ao concluir o ensino médio, continuar estudando e trabalhando, possuem 48,6% de probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB.

A segunda árvore de decisão apresenta 7 nós sendo 4 deles considerados nós terminais (3, 4, 5 e 6) a amostra incluída na árvore teve tamanho $n = 2278$ e analisou a proficiência dos estudantes (abaixo da média e acima da média) com significância para as seguintes variáveis: Wi-Fi em casa (sim e não) e Notebook em casa (possui e não possui).

Figure 1: Future Variable Tree.



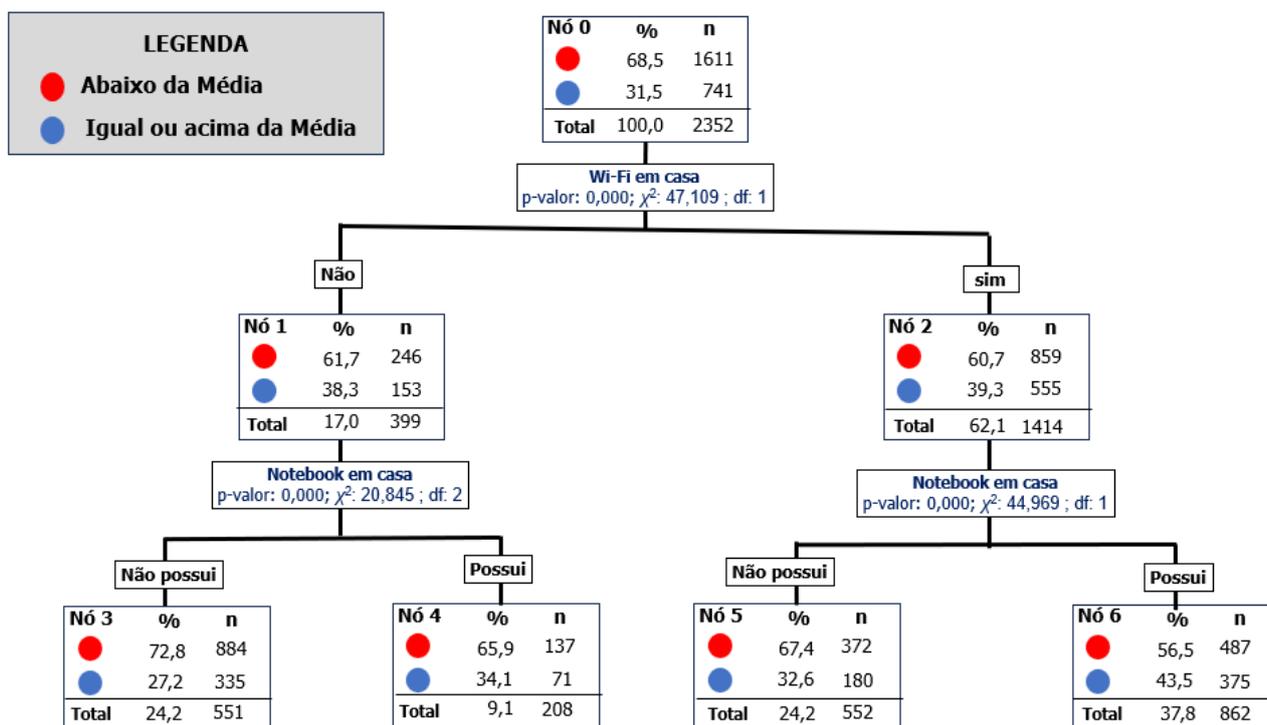
Source: from the authors (2024).

A Figura 2 apresenta os perfis de probabilidades para identificar estudantes do Amapá com mais chances de estarem abaixo da média nacional em Língua Portuguesa na prova do SAEB.

Os perfis identificados foram:

- Do nó 0 ao nó 3: estudantes que não tem Wi-Fi em casa, que não possui notebook: possuem 72,8% de probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB;
- Do nó 0 ao nó 5: estudantes que tem Wi-Fi em casa, e não possui notebook: possuem 67,4% de probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB;
- Do nó 0 ao nó 4: estudantes que não tem Wi-Fi em casa, possui notebook: possuem 65,9% de probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB;
- Do nó 0 ao nó 6: estudantes que tem Wi-Fi em casa, possui notebook: possuem 56,5% de probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB;

Figure 2: Technology Variable Tree.



Source: from the authors (2024).

A terceira árvore de decisão apresenta 14 nós sendo 8 deles considerados nós terminais (4, 7, 8, 9, 10, 11, 12 e 13) a amostra incluída na árvore teve tamanho $n = 2307$ e analisou a proficiência dos estudantes (abaixo da média e acima da média) com significância para as seguintes variáveis: raça/cor (branco, amarela, pardo, preta, indígena e não declarados), ler livros (nunca ou quase nunca, de vez em quando e sempre ou quase sempre) e ler notícias (nunca ou quase nunca, de vez em quando e sempre ou quase sempre).

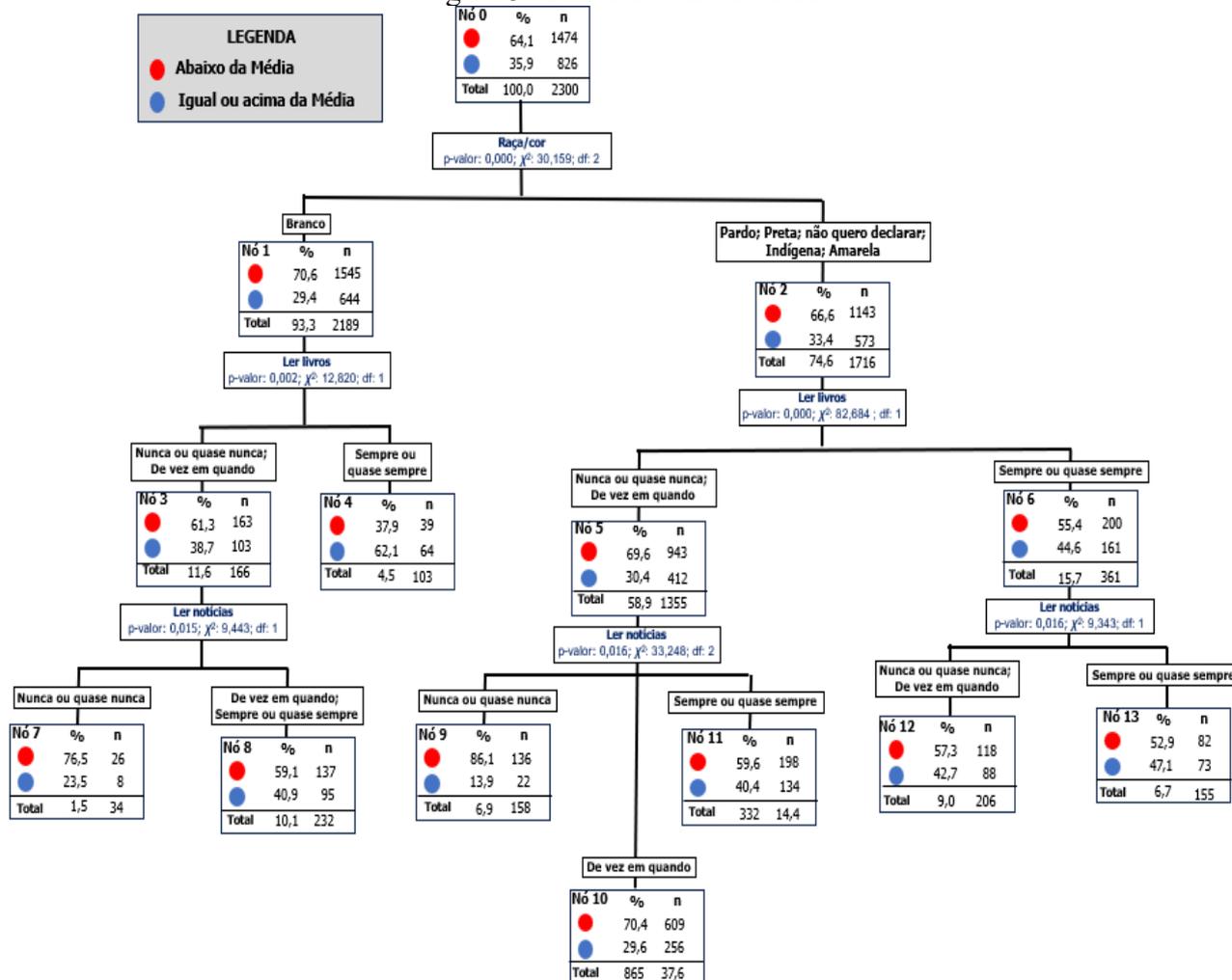
A Figura 3 apresenta os perfis de probabilidades para identificar estudantes do Amapá com mais chances de estarem abaixo da média nacional em Língua Portuguesa na prova do SAEB.

Os perfis identificados foram:

- a) Do nó 0 ao nó 9: estudantes de raça/cor pardo/amarela/preta/indígena que de vez em quando, nunca ou quase nunca ler livros e notícias: possuem 86,1% probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB;
- b) Do nó 0 ao nó 7: estudantes de raça/cor branco que de vez em quando, nunca ou quase nunca ler livros e notícias: possuem 76,5% probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB;
- c) Do nó 0 ao nó 10: estudantes de raça/cor pardo/amarela/preta/indígena que de vez em quando, nunca ou quase nunca ler livros e de vez em quando ler notícias: possuem 70,4% probabilidade de estarem abaixo da média nacional na prova de Matemática SAEB;
- d) Do nó 0 ao nó 11: estudantes de raça/cor pardo/amarela/preta/indígena que de vez em quando, nunca ou quase nunca ler livros e sempre ou quase sempre ler notícias: possuem 59,6% probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB;

- e) Do nó 0 ao nó 8: estudantes de raça/cor branco que de vez em quando, nunca ou quase nunca ler livros e de vez em quando, sempre ou quase sempre ler notícias: possuem 59,1% probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB;
- f) Do nó 0 ao nó 12: estudantes de raça/cor pardo/amarela/preta/indígena que sempre ou quase sempre ler livros e nunca ou quase nunca, de vez em quando ler notícia: possuem 57,3% probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB.
- g) Do nó 0 ao nó 13: estudantes de raça/cor pardo/amarela/preta/indígena que sempre ou quase sempre ler livros e notícia: possuem 52,9% probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB.
- h) Do nó 0 ao nó 4: estudantes de raça/cor branco que sempre ou quase sempre ler livros: possuem 37,9% probabilidade de estarem abaixo da média nacional na prova de Língua Portuguesa SAEB.

Figure 3: Cultural Variable Tree.



Source: from the authors (2024).

A análise da qualidade de classificação do modelo CHAID ocorre, segundo Pestana e Gageiro (2009), com base nos ganhos percentuais de classificação da categoria de interesse por nó

(tabela de ganhos por nós) e também nos ganhos percentuais segundo os percentis (tabela de ganhos por percentis).

A Tabela 1 seguinte mostra que na árvore 1, o nó com maior contribuição na classificação de alunos abaixo da média na nota de Português é o nó 6 que classifica 75,7% em relação a amostra total de alunos que foram categorizados com nota abaixo da média. Já na árvore 2, o nó com maior contribuição na classificação de alunos é o nó 6 que classifica 32,8% dos alunos como abaixo da média em Português. Para a árvore 3, o nó com maior contribuição na classificação de alunos abaixo da média em Português é o nó 10 com 41,3% de classificação.

Table 1: Percentage gains of the "Below Average" category by node.

Árvore	Nó	Amostra Total Nó		Ganho Categoria Alvo		Resposta	Índice
		N	Porcentagem	N	Porcentagem		
Árvore 1	8	57	2,4%	51	3,3%	89,5%	139,1%
	4	35	1,5%	26	1,7%	74,3%	115,5%
	7	149	6,3%	128	8,4%	85,9%	133,5%
	6	1798	75,9%	1153	75,7%	64,1%	99,7%
	3	284	12,0%	138	9,1%	48,6%	75,5%
	5	46	1,9%	28	1,8%	60,9%	94,6%
Árvore 2	3	551	24,2%	401	27,0%	72,8%	111,8%
	5	552	24,2%	372	25,1%	67,4%	103,5%
	4	208	9,1%	137	9,2%	65,9%	101,2%
	6	862	37,8%	487	32,8%	56,5%	86,8%
	9	158	6,9%	136	9,2%	86,1%	134,3%
	7	34	1,5%	26	1,8%	76,5%	119,3%
Árvore 3	10	865	37,6%	609	41,3%	70,4%	109,9%
	11	332	14,4%	198	13,4%	59,6%	93,1%
	8	232	10,1%	137	9,3%	59,1%	92,1%
	12	206	9,0%	118	8,0%	57,3%	89,4%
	13	155	6,7%	82	5,6%	52,9%	82,5%
	4	103	4,5%	39	2,6%	37,9%	59,1%

Source: from the authors (2024).

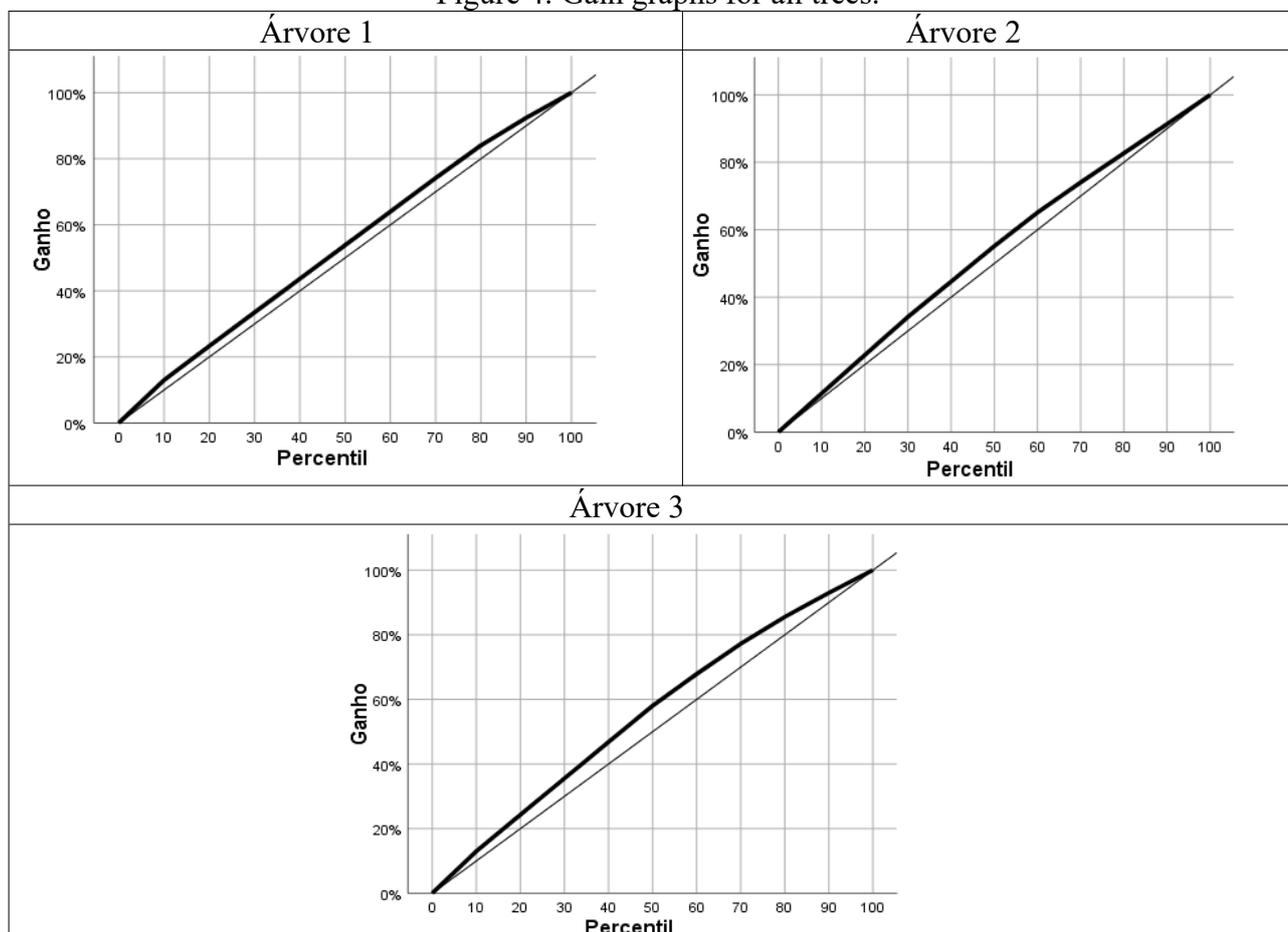
Na árvore 1, observa-se também que, o nó 8 possui 51 dos 57 alunos classificados com notas abaixo da média, totalizando 89,5% dentro do nó, sendo este o nó com maior percentual de classificação quando avaliado somente o nó. Na árvore 2, o nó 3 possui 401 dos 551 alunos classificados com notas abaixo da média, totalizando 72,9% dentro do nó, sendo este o nó com maior percentual de classificação quando avaliado somente o nó. Na árvore 3, o nó 9 possui 136 dos

158 alunos classificados com notas abaixo da média, totalizando 86,1% dentro do nó, sendo este o nó com maior percentual de classificação quando avaliado somente o nó.

A análise dos ganhos percentuais considerando os percentis ocorre de modo análogo ao apresentado na tabela anterior, levando-se em conta que os nós terminais das árvores estão em função dos percentis. A Tabela 2 seguinte destaca que, na árvore 1, 10% da amostra (distribuída nos nós 8, 4 e parte de 7) contém 85,0% de estudantes classificados como abaixo da média nacional. Na árvore 2, 10% da amostra (distribuída no nó 3) contém 77,0% de estudantes classificados como abaixo da média nacional. Na árvore 3, 10% da amostra (distribuída nos nós 9 e 7) contém 81,4% dos estudantes classificados como abaixo da média nacional.

A Figura 4 apresenta os gráficos de percentis e ganhos (Pestana; Gageiro, 2009), onde cada gráfico representa uma árvore gerada pelo modelo. A função desses gráficos é indicar a capacidade do modelo estatístico em explicar a classificação. Segundo Pestana e Gageiro (2009), quanto mais a curva estiver acima da linha diagonal, mais eficiente será a previsão do modelo. Como pode ser observado abaixo, todos os modelos demonstraram minimamente capacidade em prever o desempenho dos alunos abaixo da média.

Figure 4: Gain graphs for all trees.



Source: from the authors (2024).

Table 2: Gains of the nodes based on percentiles.

Árvores	Percentil	Nós	Ganhos				
			N	N	Porcentagem	Resposta	Índice
Árvore 1	10	8; 4; 7	237	201	13,2%	85,0%	132,2%
	20	7; 6	474	354	23,2%	74,8%	116,2%
	30	6	711	506	33,2%	71,2%	110,7%
	40	6	948	658	43,2%	69,5%	108,0%
	50	6	1185	810	53,2%	68,4%	106,3%
	60	6	1421	962	63,1%	67,7%	105,2%
	70	6	1658	1114	73,1%	67,2%	104,4%
	80	6	1895	1266	83,1%	66,8%	103,8%
	90	6; 3	2132	1403	92,1%	65,8%	102,3%
	100	3; 5	2369	1524	100,0%	64,3%	100,0%
Árvore 2	10	3	228	175	11,8%	77,0%	118,3%
	20	3	456	341	23,0%	74,9%	115,0%
	30	3; 5	683	505	34,1%	74,0%	113,6%
	40	5	911	659	44,4%	72,3%	111,1%
	50	5	1139	813	54,8%	71,3%	109,6%
	60	5; 4	1367	964	65,0%	70,5%	108,3%
	70	4; 6	1595	1097	74,0%	68,8%	105,7%
	80	6	1822	1226	82,6%	67,3%	103,3%
	90	6	2050	1354	91,3%	66,1%	101,5%
	100	6	2278	1483	100,0%	65,1%	100,0%
Árvore 3	10	9; 7	230	187	12,7%	81,4%	127,0%
	20	10	460	344	23,3%	74,7%	116,6%
	30	10	690	506	34,3%	73,3%	114,4%
	40	10	920	668	45,3%	72,6%	113,2%
	50	10	1150	830	56,3%	72,1%	112,6%
	60	10; 11	1380	975	66,1%	70,6%	110,2%
	70	11; 8	1610	1102	74,7%	68,4%	106,8%
	80	8; 12	1840	1237	83,9%	67,2%	104,9%
	90	12; 13	2070	1368	92,8%	66,1%	103,1%
	100	13; 4	2300	1474	100,0%	64,1%	100,0%

Source: from the authors (2024).

De acordo com Pestana e Gageiro (2009), o modelo de árvore de decisão CHAID apresenta algum risco de classificação incorreta da amostra total de estudantes. Esse risco é medido em porcentagem e está indicado na coluna de Intervalo de Confiança (IC) da Tabela 3. Desta forma, para a árvore 1, o modelo tem o risco de classificar incorretamente cerca de 34% a 38% dos alunos que compõem a amostra total. A mesma interpretação deve-se fazer para os demais intervalos das demais árvores.

Table 3: Risk Estimates and Confidence Intervals for Prediction Errors.

Árvore	Risco	Erro	%	IC
1	0,361	0,010	36,1%	[34%; 38%]
2	0,349	0,010	34,9%	[32%; 36%]
3	0,352	0,010	35,2%	[33%; 37%]

Source: from the authors (2024).

Neste estudo, observou-se que os perfis de estudantes com maior probabilidade de apresentarem notas abaixo da média foram de alunos de raça/cor preta, pardo, amarela e indígena, pouco acesso às tecnologias (Wi-Fi e notebook) e que não leem livros ou notícias.

Na literatura comparada, não foram identificados estudos que analisam perfis de alunos com estimativas de probabilidades abaixo da média ou acima da média na prova do SAEB, mas existem estudos com objetivos semelhantes e empregos de outros métodos para identificar perfis de vulnerabilidades, como no estudo de Diaz (2012) que estudou desigualdades entre estudantes de escolas públicas e privadas do Ensino Médio, com base nas notas de Português e Matemática do SAEB.

O estudo de Diaz (2012) utilizou um índice de igualdade de oportunidades e seus resultados mostraram que as desigualdades de oportunidades entre estudantes de escolas públicas e privadas foram menores em Língua Portuguesa, cuja diferença se situou em torno de 16% da desigualdade total e, em Matemática essa diferença foi próxima de 24%.

Já nos estudos de Gusmão e Amorim (2022) os autores objetivaram examinar a relação entre o desempenho escolar e os indicadores de desigualdade social e econômica (PIB, IDH, Índice de Gini) das Unidades Federativas do Brasil utilizando as notas do SAEB em Língua Portuguesa e Matemática de alunos de Ensino Médio do Brasil. Os autores concluíram que, existe correlação positiva entre os PIB estaduais e o rendimento escolar em Língua Portuguesa e em Matemática, quando analisado o recorte temporal de 2013 a 2015, inferindo que, estados com elevado PIB possui as notas mais elevadas em Língua Portuguesa e em Matemática. O mesmo também foi observado para o Índice de Desenvolvimento Humano, pois estados com melhores IDH também são aqueles que possuem os melhores rendimento de Língua Portuguesa e Matemática no SAEB.

Nesta ótica, o estudo ora apresentado contribui com a literatura e ressaltou a importância do SAEB como um sistema de grande importância de avaliação educacional nacional. No entanto, é fundamental transformar seus resultados em medidas políticas concretas, baseadas nos diagnósticos fornecidos por essa e por demais pesquisas que visem expor as desigualdades dos estudantes de Ensino Médio em relação ao desempenho em Língua Portuguesa e em Matemática. Isso garantiria a utilização efetiva dos resultados do SAEB, contribuindo assim para o aprimoramento do IDEB e seu impacto social.

Conclusões

O presente estudo teve como objetivo geral analisar o desempenho escolar dos alunos do Ensino Médio do Amapá na disciplina de Língua Portuguesa, com foco na identificação dos perfis de estudantes com maior probabilidade de apresentar notas abaixo da média nacional do SAEB referente ao ano de 2019. Para atingir esse propósito, foi realizado o estudo dos microdados disponibilizados pelo INEP para o referido ano e selecionadas as informações pertinentes aos alunos do Amapá. Posteriormente, foram conduzidas análises estatísticas descritivas e aplicado um algoritmo de árvore de decisão, possibilitando a identificação dos fatores associados ao desempenho acadêmico desses estudantes.

Com desafios como falta de recursos educacionais e desigualdades socioeconômicas, entender as causas do baixo desempenho escolar permite criar políticas educacionais mais eficazes. Isso não só ajuda os alunos a superar obstáculos específicos, mas também promove um desenvolvimento mais equitativo e fortalece as políticas públicas do estado ao melhor capacitar a população para os desafios socioeconômicos do futuro. As informações e resultados gerados por esta pesquisa podem servir de base para a formulação de políticas educacionais mais direcionadas e eficazes, adaptadas às necessidades específicas da região.

Referências

- Coelho, G. R. B; Lopes, T. B; Costa, A. B. (2022). **Comparativo nacional do desempenho da rede federal de ensino no SAEB de 2013 e 2015**. Research, Society and Development, v. 11, n. 4, e38511427703. Disponível em: <https://doi.org/10.33448/rsd-v11i4.27703>
- DIAZ, M. D. M. (2012). (Des)igualdades de oportunidades no ensino médio brasileiro: **Escolas públicas e privadas**. *Economia*, 13(3a), 553-568. Disponível em: https://www.anpec.org.br/revista/vol13/vol13n3ap553_568.pdf. Acesso em: 01 out. 2024.
- GUSMÃO, F. A. F.; AMORIM, S. S. **Desigualdade educacional no ensino médio brasileiro**. *Educação*, v. 47, n. 1, p. e108, 2022. Disponível em: https://www.researchgate.net/profile/Fabio-Alexandre-Ferreira-Gusmao/publication/365703210_Desigualdade_educacional_no_ensino_medio_brasileiro/links/637f2a281766b34c544f182a/Desigualdade-educacional-no-ensino-medio-brasileiro.pdf. Acesso em: 01 out. 2024.
- INEP. **Microdados do INEP, 2024**. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados>.
- INEP. **Resultados do SAEB, 2019**. Disponível em: <https://undime.org.br/noticia/16-09-2020-09-24-inep-divulga-resultados-do-saeb-2019>.
- LAPEI UFG. **Uso de dados na pesquisa: qual a diferença entre dados primários e secundários?**. 2024. Disponível em: <https://lapei.face.ufg.br/p/43695-17-uso-de-dados-na-pesquisa-qual-a-diferenca-entre-dados-primarios-e-secundarios>.

Pestana, M. H; Gageiro, J. N. (2009). **Análise categórica, árvores de decisão e análise de conteúdo em ciências sociais e da saúde com o SPSS**. Lisboa; Porto: Lidel, D.L. 2009. - XVIII, 551, [7] p. : il. ; 24 cm. - (Estatística). - Bibliografia, p. 543-546. - ISBN 978-972-757-585-5.

Silva, E. L. da.; Menezes, E. M. **Metodologia de pesquisa e elaboração de teses e dissertações, 2005**. 4. ed. rev. atual. Universidade Federal de Santa Catarina – UFSC. Florianópolis 2005.

Silva, I. V; Silva, M. T; Martins, S. (2018). **Análise do desempenho escolar na avaliação SAEB 2015**. In XXXVIII Encontro Nacional de Engenharia de Produção "A Engenharia de Produção e suas contribuições para o desenvolvimento do Brasil". Maceió, Alagoas, Brasil. Egenep.