UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO

KIMBERLY TENAN RIBEIRO

O uso de árvores de regressão em grande base de dados: aplicação para o caso do desempenho de escolas brasileiras no ensino fundamental

São Paulo
2024

KIMBERLY TENAN RIBEIRO

O uso de árvores de regressão em grande base de dados: aplicação para o caso do desempenho de escolas brasileiras no ensino fundamental

Versão original

Trabalho de Formatura apresentado à Escola Politécnica da Universidade de São Paulo para obtenção do Diploma de Engenheiro de Produção

Área de concentração: Engenharia de Produção.

Orientadora: Profa. Dra. Linda Lee Ho

São Paulo
2024

FICHA CATALOGRÁFICA

Ribeiro, Kimberly Tenan

O uso de árvores de regressão em grande base de dados: aplicação para o caso do desempenho de escolas brasileiras no ensino fundamental/ K. T. Ribeiro - São Paulo, 2024.

115p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

1.Ciência de dados. 2.Árvores de regressão. 3.Desempenho escolar. 4.*Big Data*. 5.Educação brasileira. 6.Modelagem com *rpart*. I. Ribeiro, Kimberly Tenan. II.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção III.Título.

AGRADECIMENTOS

A entrega desse trabalho é muito importante para mim, não é só a materialização das conquistas dos cinco anos da graduação, mas também a consolidação do sucesso que os demais anos de estudo, esforço e dedicação ajudaram a construir. Essa entrega é tão minha quanto de todos que me incentivaram e estiveram comigo durante esse ciclo tão desafiador. Meus principais agradecimentos:

Aos meus pais, Sandra e Osvaldo, que lutaram por mim e comigo durante todos os dias da minha vida, colocaram minha educação e bem estar como prioridades inegociáveis e nunca deixaram de acreditar no potencial de suas filhas. Agradeço por todas as idas e vindas dos pontos de ônibus, pelos jantares depois de um dia longo e pela ansiedade que passaram comigo nas vésperas de cada vestibular, prova, entrega e entrevista. Sou grata por todo carinho que vocês expressam por mim, desde os gestos comuns do dia a dia, até os sacrifícios para as grandes conquistas. Muito disso é de vocês também.

À minha irmã, Evelyn, que abriu caminho e serviu de modelo para mim em diversas situações. Foi amiga, conselheira e orientadora em todos assuntos da faculdade e em todas as grandes decisões da minha vida. Assim como meus pais, zela por mim, e é uma grande inspiração ao desafiar-se, estudar e tornar-se uma pessoa ainda melhor a cada dia. Te admiro, irmã.

Ao meu amor, Carlos Eduardo, que é também meu melhor amigo dos últimos cinco anos. Você está comigo a cada passo, descoberta e anseio, é quem ouve atento a cada detalhe de todo dia e me aconselha sobre praticamente tudo. Você acreditou em mim antes de eu aprender a fazê-lo e vê de perto a construção de quem eu sou, comemora meus passos assim como os seus e faz eu me sentir a pessoa mais capaz do mundo. Obrigada por ser meu grande parceiro.

A todos amigos que fiz durante essa jornada, que compartilharam os dias de aula, de festas e de tristezas como uma família. Aos do Millenium - Carol, Júlia, Rodrigo, Babs, Posztos - aos do Liceu - Bia Lessa, Tatá, Bruninho - e da Poli - Sara, Camilla, Fernanda, Bia Calil, Producats, Zunta, Soclaus, Caepers. Saibam que vocês são parte disso.

Aos demais membros da minha família, tias e primas, aos meus sogros, e, por último, ao meus *pets* Pingo e Mia, que sempre foram os companheirinhos mais doces e amorosos do mundo. Sou grata por ter uma rede de apoio tão maravilhosa ao meu lado, que me possibilitou chegar onde estou e ter confiança para continuar minha trajetória.

Por fim, um agradecimento especial à Linda Lee Ho, minha orientadora, que me auxiliou durante esse ano ao me direcionar às melhores decisões e aprendizados na elaboração deste trabalho, e à Escola Politécnica pelo conhecimento transmitido e adquirido.

RESUMO

RIBEIRO, Kimberly Tenan. **O uso de árvores de regressão em grande base de dados: aplicação para o caso do desempenho de escolas brasileiras no ensino fundamental.** 2024. Trabalho de formatura - Escola Politécnica da Universidade de São Paulo, São Paulo, 2024.

O presente estudo aplica árvores de regressão na análise de grandes bases de dados com enfoque específico no caso do desempenho das escolas brasileiras do ensino fundamental. O trabalho teve como objetivo principal identificar os fatores críticos que influenciam o desempenho acadêmico dentre variáveis socioeconômicas, geográficas, estruturais e de gestão escolar. Para isso, manipulou-se o banco de dados do Sistema de Avaliação do Ensino Básico (SAEB) de 2019. Após a revisão literária, foram criadas árvores de regressão através da metodologia proposta: partiu-se da obtenção e tratamento dos dados, realização da análise exploratória das variáveis e encaminhou-se até os ciclos de modelagem e avaliação de precisão para cada modelo gerado (em linguagem R e biblioteca *rpart*). A interpretação das árvores revelou as variáveis mais relevantes para o desempenho educacional.

Esses resultados podem embasar a formulação de políticas públicas para aprimoramento da gestão educacional no Brasil. Com esse método, altamente replicável para outras grandes bases de dados, destaca-se o papel da ciência de dados para a resolução de problemas complexos e a tomada de decisões baseadas em dados, tanto no contexto da engenharia quanto de cunho social e para diversas áreas do conhecimento.

Palavras chave: Ciência de dados. Árvores de regressão. Desempenho escolar. *Big Data*. Educação brasileira. Modelagem com *rpart*.

ABSTRACT

RIBEIRO, Kimberly Tenan. **The use of regression trees in large datasets: application to the case of brazilian elementary school performance.** 2024. Trabalho de formatura - Escola Politécnica da Universidade de São Paulo, São Paulo, 2024.

The present study applies regression trees to the analysis of large datasets, with a specific focus on the performance of Brazilian elementary schools. The main objective was to identify critical factors influencing academic performance among socioeconomic, geographic, structural, and school management variables. To achieve this, the 2019 Basic Education Assessment System (SAEB) database was manipulated. Following a literature review, regression trees were created using the proposed methodology. This involved data acquisition and processing, exploratory analysis of the variables, and iterative cycles of modeling and accuracy evaluation for each generated model (using the R programming language and the *rpart* library). The interpretation of the trees revealed the most relevant variables for educational performance.

These findings can support the formulation of public policies to improve educational management in Brazil. This method, which is highly replicable for other large datasets, highlights the role of data science in solving complex problems and enabling data-driven decision-making, both in engineering and social contexts and across various knowledge domains.

Keywords: Data science. Regression trees. School performance. *Big Data*. Brazilian education. Modeling with *rpart*.

LISTA DE FIGURAS

LISTA DE TABELAS

Sumário

1 INTRODUÇÃO

Neste capítulo, será introduzido o tema central do trabalho. O contexto da ciência de dados e sua aplicação na análise de grandes volumes de dados *(Big Data)* serão abordados, assim como o panorama atual da educação no Brasil e os principais indicadores e avaliações de desempenho escolar. Seguirá então a apresentação do caso proposto para análise de dados no campo da educação brasileira e as definições do problema central, objetivo e escopo do estudo.

1.1 Contexto da ciência e análise de dados

A necessidade e o valor gerado pela análise de dados e uso de ferramentas estatísticas têm sido evidentes nas últimas décadas, intensificando-se com a rápida evolução tecnológica e a grande quantidade de dados gerados no monitoramento de informações. Com a análise desses dados, torna-se possível a extração de percepções, a previsão de comportamentos e consequentemente o embasamento de decisões estratégicas em organizações e nas políticas da sociedade e mercado. Esse movimento impulsionou o surgimento da ciência de dados, um campo interdisciplinar que combina estatística e programação para transformar grandes volumes de dados em informações acionáveis. O processo de ciência de dados inclui a coleta, limpeza, processamento e análise de dados, utilizando linguagens de programação como Python e R.

Assim, o conceito que melhor sintetiza a ideia apresentada é o *Big Data*. *Big Data* refere-se ao conjunto massivo de dados complexos e variados, gerados em alta velocidade e que exigem tecnologias avançadas para armazenamento, processamento e análise. Esses dados vêm de fontes diversas, como redes sociais, sensores, transações comerciais, pesquisas do tipo censo, entre outros.

A importância do *Big Data*, como apontado por McAfee e Brynjolfsson (2012)[1], está na sua capacidade de fornecer *insights* valiosos que antes eram inatingíveis devido à limitação dos métodos computacionais. A análise de *Big Data* permite identificar padrões e tendências, prever comportamentos e embasar a tomada de decisões. Manyika et al. (2011) [2] destacam que essa tecnologia de processamento tem o potencial de transformar diversos setores, gerando valor ao melhorar a eficiência operacional e impulsionar a inovação.

Nesse contexto, a definição de *Big Data* é utilizada para sintetizar o objeto de estudo e manipulação do presente trabalho: Busca-se compreender e implementar uma das possíveis maneiras de operar grandes volumes de dados a fim de gerar embasamento para medidas práticas e tomada de decisões. Além disso, por entender-se que a ciência de dados possui aplicações em diversas áreas, como negócios, engenharia e ciências sociais ao ajudar a resolver problemas complexos a partir de dados, para esse trabalho, será realizado um estudo sobre o caso do desempenho da educação brasileira. Nele, serão utilizadas bases de dados oficiais e ferramentas estatísticas, como análise exploratória e árvores de regressão - as quais serão definidas posteriormente - como insumos do estudo.

1.2 Contexto para o caso da educação brasileira

De acordo com os resultados obtidos no Programa Internacional de Avaliação de Estudantes (Pisa[1]) de 2022 e divulgados pela Organização para Cooperação e Desenvolvimento Econômico (OCDE), no Brasil, pelo menos metade dos estudantes de 15 anos não atingiu padrão mínimo de aprendizado - considerado como o nível limite em que os jovens podem exercer plenamente sua cidadania. Ao comparar os resultados da avaliação em um ranking decrescente com outras 80 nações, o país ocupou as posições 53ª em Leitura, 61ª em Ciências e 65ª em Matemática, os três domínios de conhecimentos avaliados. Desde 2009, os resultados são estáveis para as três disciplinas, mas encontram-se abaixo da média geral. (INEP, 2023) [3].

Já no âmbito nacional, os resultados do Sistema de Avaliação do Ensino Básico - SAEB (conjunto de avaliações aplicadas pelo Ministério da Educação do Brasil para medir a qualidade do ensino nas escolas brasileiras) também são estáveis, mas aquém do desejado. Para essa avaliação, que ocorre a cada dois anos e abrange o desempenho da educação básica (ensino infantil, ensino fundamental e ensino médio), as médias gerais históricas oscilam entre os níveis mais baixos de uma escala de proficiência, nos quais os estudantes também são considerados deficitários em aprendizagem.

Além do Pisa e do SAEB, temos o Índice de Desenvolvimento da Educação Básica - IDEB (criado em 2007 e que combina o fluxo escolar e as médias de desempenho nas

[1]Segundo o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), o Pisa é um estudo comparativo internacional sobre o desempenho dos estudantes na faixa etária dos 15 anos. Seus resultados fornecem dados valiosos para melhorar políticas educacionais. No Brasil, o Inep coordena todas as etapas do processo de avaliação e análise dos resultado; o estudo é realizado a cada três anos.

avaliações) com o qual também é possível medir a qualidade da educação no Brasil. Esse índice varia de 0 a 10, tem o valor 6 como meta – com a intenção de ser comparável aos padrões de países desenvolvidos – e é um bom orientador para a melhoria dos sistemas educacionais, ajudando a guiar políticas públicas. No entanto, para seu último resultado, em 2021, seu valor foi de 5 para o ensino fundamental e 3,9 no ensino médio, valores distantes da meta proposta.

Utilizando essas ferramentas de avaliação de aprendizado para um diagnóstico geral, o cenário da educação brasileira atual fica claro: o país está longe da excelência. É notável a defasagem educacional duradoura e histórica, já que índices e avaliações continuam estagnados em baixos patamares por anos. A partir desse contexto, o problema focado e a forma de abordagens propostos nesse trabalho serão abordados.

1.3 Definição do problema

Tendo em vista o contexto apresentado, nota-se que o Brasil têm sido um país de destaque negativo para o setor educacional, dado que o desempenho de escolas e alunos brasileiros frente aos de outros países está entre os piores posicionados. Ainda na própria linha temporal brasileira, os indicadores de desempenho não têm melhorado significantemente e encontram-se estagnados.

Um dos primeiros passos para transformar essa situação é estabelecer uma tentativa de explicar o seguinte problema: **"O que tem influenciado o desempenho brasileiro na educação?"** Ao responder essa questão, será possível entender quais aspectos, além das questões de desempenho individuais e características de intelectualidade própria aos alunos, possivelmente têm levado escolas a desempenharem mal e gerar baixos índices educacionais como consequência. Aspectos como características físicas, estruturais e de gestão de recursos de escolas, poderiam explicar a variabilidade de desempenho de escolas, e isso será investigado neste trabalho. Munidas dessas informações, políticas públicas podem aumentar sua assertividade e eficiência.

Além disso, tal questão pode ser **respondida com o auxílio de técnicas de ciência de dados**, que descrevem a relação entre variáveis explicativas (como infraestrutura escolar, gestão de recursos e outras características institucionais) e o desempenho educacional. Com a aplicação das ferramentas estatísticas, será possível identificar os fatores mais relevantes que influenciam o desempenho, dividindo os dados em grupos mais

homogêneos e interpretáveis, e relacionando uma questão social com métodos matemáticos de análise de dados.

1.4 Objetivo e escopo do estudo

Define-se como um dos objetivos principais do estudo o entendimento do desempenho das escolas brasileiras, a fim de indicar as questões geográficas, de gestão e de recursos que, quando combinadas, levariam uma instituição a melhores e piores desempenhos. Também é objetivo a proposição de um método de análise de dados pautado na ciência de dados e *Big Data* para guiar e responder questões complexas do tipo.

Através do emprego da técnica de Árvores de Regressão (as quais serão discutidas em profundidade nas próximas seções), o objetivo é identificar quais características e seus respectivos níveis são capazes de explicar variabilidade entre as notas de desempenho escolar.

Ao utilizar as bases de dados advindas do SAEB de 2019 - que consistem em um conjunto de informações do contexto das avaliações aplicadas nas escolas públicas em nível fundamental e médio - como insumo para o estudo, será possível obter um registro de localidade, porte, infraestrutura (dependências da escola), recursos (corpo docente, acesso livros didáticos, acessibilidade), gestão (presença de grêmio, conselho tutelar, reforço), entre outras variáveis para cada escola brasileira. Além disso, para cada escola são disponibilizadas as notas obtidas nas avaliações e a definição do seu desempenho em níveis de aprendizado.

Dessa forma, o escopo do trabalho será reunir as bases de dados do SAEB, estruturar e descrever suas informações, verificar sua consistência e utilizá-las como entrada para modelos de árvores de regressão (que serão capazes de delimitar quais características da escola estão presentes naquelas que obtiveram melhores notas). Como saída, teremos a relação das variáveis que mais distinguem as notas das escolas nas avaliações de língua portuguesa e matemática para o quinto ano do fundamental das escolas brasileiras, o primeiro nível avaliado pelo SAEB 2019 fora de fase de testagem, onde entende-se que o problema educacional começaria a ser notado e também para o qual temos informações e dados abrangentes.

1.5 Relevância e motivação

O desempenho escolar é um indicador fundamental da qualidade do sistema educacional em um país, assim, o estudo justifica sua relevância ao contribuir para uma melhor compreensão desse indicador no Brasil e ao propor uma abordagem de análise de dados que visa otimizar decisões. A ciência de dados, com a aplicação de técnicas avançadas de análise e modelagem, oferece ferramentas poderosas para transformar grandes volumes de dados em informações acionáveis, não apenas para questões educacionais, mas também para desafios sociais mais amplos.

Esse estudo pode também contribuir para a engenharia de produção, fornecer informações sobre como otimizar sistemas educacionais, identificar pontos de intervenção para melhorar escolas e até mesmo para entender melhor a relação entre educação e o enriquecimento do capital humano por meio de melhorias na qualidade da educação.

Para Veloso (2022) [4]

> A melhoria da educação é fundamental para o desenvolvimento econômico, na medida em que aumenta o capital humano dos trabalhadores e facilita a criação e absorção de novas tecnologias. Além disso, contribui para uma inserção mais produtiva da população no mercado de trabalho, o que se manifesta de diversas formas, como aumento do salário e maior probabilidade de obtenção de um emprego formal.

Também para Schwartzman e Castro (2013)[5]

> A necessidade de melhor qualificação dos recursos humanos é um requisito da economia e uma aspiração da população, que sabe que as pessoas mais educadas conseguem melhores empregos e melhores rendas. Quando recursos humanos de qualidade escasseiam e o sistema educacional não responde, a economia tende a se ajustar a esta situação, desenvolvendo atividades baseadas em trabalho de baixa qualificação.

Dessa forma, reforça-se a ideia de que a melhoria educacional seja vital para o desenvolvimento econômico e que deve ser objeto de análise. Essa melhoria também é responsável por aumentar o capital humano e sua qualificação, dado que a educação de qualidade leva a obtenção de melhores empregos e rendas para os futuros trabalhadores que hoje são estudantes. Em suma, a proposta de pesquisa sobre a relevância das variáveis do SAEB para a compreensão do desempenho escolar no Brasil através das árvores de regressão tem o potencial de fornecer elucidações para a melhoria de políticas públicas e práticas educacionais – e consequente enriquecimento de mão-de-obra – e demonstra

a aplicação prática de métodos estatísticos e modelagem em um contexto significativo para a sociedade, sublinhando a importância de utilizar dados para otimizar decisões e promover mudanças efetivas.

1.6 Estrutura do trabalho

Este estudo será desenvolvidos em 6 capítulos. O primeiro capítulo, já apresentado, contextualiza o tema e os objetivos do trabalho. No Capítulo 2, será descrita a revisão literária sobre o método do Sistema Nacional de Avaliação da Educação Básica (SAEB) e suas bases de dados, com foco para aquelas utilizadas na modelagem. O modelo estatístico utilizado, de Árvores de Regressão, também será abordado nesse capítulo, assim como suas formas de ajuste e medição de precisão.

No Capítulo 3, a metodologia empregada para o estudo do desempenho educacional será abordada, passando pela forma de obtenção dos dados, a estruturação e tratamento das bases do SAEB, a forma pela qual a análise descritiva das variáveis envolvidas foi realizada e suas premissas, e por fim o plano de modelagem das árvores de regressão. No Capítulo 4, os resultados da estruturação da base e da análise exploratória com a estatística descritiva serão apresentados. Para o Capítulo 5, os resultados da modelagem final das árvores serão discutidos, finalizando cada etapa com os dados finais obtidos no processo.

Por fim, no Capítulo 6, apresentam-se as conclusões do trabalho, discutindo e retomando o método empregado e resultados obtidos, e também as lacunas e formas de melhorar essas informações, ou ainda de tornar factível o uso delas em maiores projetos educacionais e políticas públicas.

2 REVISÃO BIBLIOGRÁFICA

Para o desenvolvimento do presente trabalho com os objetivos apresentados anteriormente, é necessária a revisão e entendimento dos conceitos utilizados no estudo. Neste capítulo será dado o contexto e composição do Sistema Nacional de Avaliação da Educação Básica, e a composição das bases geradas por tal Sistema. Também serão comentadas as fases de interpretação de dados com análises exploratória e as definições e aplicações do algoritmo de Árvores de Regressão que será utilizado, além da justificativa de ter optado por seu uso frente a outros métodos de análise de dados.

2.1 Sistema Nacional de Avaliação da Educação Básica - SAEB

2.1.1 História da avaliação, periodicidade e funcionamento

De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE) [6] e o - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) [3], o SAEB foi introduzido no final dos anos 80 e sua primeira aplicação ocorreu em 1990. Inicialmente, o foco estava em avaliar a qualidade da educação básica por meio de uma amostra de escolas públicas. Em 1995, o SAEB passou por uma reformulação metodológica significativa, foi adotada a Teoria de Resposta ao Item (TRI), que possibilitou comparações mais precisas ao longo do tempo e a inclusão de dados contextuais.

A edição de 1997 introduziu matrizes de referência próprias para a construção dos testes, enquanto em 1999 foram incluídos testes de ciências humanas. Em 2001, o foco se restringiu a Língua Portuguesa e Matemática, e a avaliação de ciências humanas foi temporariamente removida. A partir de 2005, o SAEB foi reestruturado em duas avaliações distintas: a Avaliação Nacional da Educação Básica (ANEB) e a Avaliação Nacional do Rendimento Escolar (ANRESC), mais conhecida como Prova Brasil. Essa reestruturação permitiu uma avaliação censitária em escolas públicas de determinados níveis e uma avaliação amostral em escolas privadas.

Em 2007, o SAEB começou a alimentar o Índice de Desenvolvimento da Educação Básica (Ideb), que combina os resultados das avaliações com dados sobre taxas de aprovação e reprovação. Em 2013 e 2014, a Avaliação Nacional da Alfabetização (ANA) foi integrada ao SAEB, que foca na alfabetização no 3º ano do ensino fundamental e amplia a avaliação para ciências humanas e ciências da natureza. Em 2019, o SAEB passou por

uma reestruturação para se alinhar com a Base Nacional Comum Curricular (BNCC), substituindo as siglas ANEB e ANRESC pela nomenclatura única SAEB, abrangendo novas matrizes de referência.

Sobre sua periodicidade, O SAEB é realizado de forma bienal, com edições em anos alternados desde seu início. A periodicidade garante que a avaliação possa refletir mudanças e tendências ao longo do tempo. As avaliações são realizadas no 5º e 9º ano do ensino fundamental e na 3ª série do ensino médio, e recentemente, também no 2º ano do ensino fundamental e na educação infantil. O SAEB utiliza diversas metodologias para garantir a precisão e a relevância dos dados, que incluem as matrizes de referência, testes padronizados, questionários de contexto, Teoria de Resposta ao Item (TRI) e escalas de proficiência. Isso permite uma análise detalhada do desempenho dos alunos e das condições educacionais, além de oferecer subsídios para a formulação e monitoramento de políticas públicas.

Para o Ensino Fundamental, a avaliação do SAEB é censitária nas escolas públicas. Isso significa que todas as escolas públicas que atendem aos critérios estabelecidos são incluídas na avaliação, sem necessidade de amostragem. Esses critérios, passíveis de exclusão das escolas são: escolas com menos de dez alunos matriculados nas séries avaliadas; escolas ou turmas que atendem apenas estudantes da Educação especial; escolas indígenas onde não se fala o português; turmas da Educação de Jovens e Adultos (EJA), de magistério e multisseriadas. Nas escolas privadas, a avaliação é feita de forma amostral, utilizam-se as técnicas de amostragem probabilística para garantir a representatividade dos resultados.

Esse estudo utilizará os dados de desempenho das escolas públicas.

2.1.2 Métricas de desempenho, níveis e faixas

O Sistema de Avaliação da Educação Básica (SAEB) emprega matrizes de referência que funcionam como guias para a construção de testes padronizados, elaboradas com o objetivo de avaliar competências e habilidades específicas, com base em currículos estaduais, municipais e na Base Nacional Comum Curricular (BNCC). Desde 2019, há uma transição gradual das matrizes anteriores, datadas de 2001, para aquelas alinhadas à BNCC, o que visa garantir maior uniformidade entre o que é ensinado nas escolas e o que é avaliado nos testes.

As matrizes de referência do SAEB são essenciais para definir o que será afe-

rido nos exames de Língua Portuguesa, Matemática, Ciências da Natureza e Ciências Humanas. As habilidades descritas nas matrizes são a base para a elaboração dos itens das provas, que buscam medir o desenvolvimento cognitivo dos estudantes em diferentes etapas da educação básica. As matrizes podem sem encontradas no site do INEP[7], e são baixadas em conjunto com o restante dos microdados[8], sendo, para o ensino fundamental, estruturadas em tópicos ou temas e respectivos descritores que indicam as competências e habilidades de Língua Portuguesa e Matemática a serem avaliadas.

Além das matrizes, o SAEB utiliza escalas de proficiência para medir o desempenho dos estudantes. Essas escalas são representações gráficas dos níveis de habilidade demonstrados nas provas. A Figura 1 - (Níveis e faixas de proficiência) representa a distribuição clara das faixas para o 5º ano do ensino fundamental e facilita a compreensão do tema.

Figura 1: Níveis e faixas de proficiência

Fonte: Plataforma Edu IDEB[9]

A pontuação obtida pelos estudantes em cada disciplina é organizada em faixas, que variam de níveis insuficientes a avançados. Por exemplo, para o 5º ano do Ensino Fundamental, na disciplina de Língua Portuguesa, alunos com pontuação até 149 pontos

são considerados em nível insuficiente, enquanto aqueles com 350 pontos ou mais atingem o nível avançado.

A avaliação em Matemática segue uma lógica similar, com estudantes do 5º ano sendo categorizados em diferentes níveis. Esses níveis de desempenho, divididos em faixas de pontuação, ajudam a identificar lacunas e direcionar ações pedagógicas para melhorar o aprendizado.

Essa estrutura de classificação e análise permite mensurar o aprendizado e também identificar tendências e vieses de desempenho entre diferentes grupos de alunos e contextos escolares, e contribui para uma avaliação mais detalhada das políticas públicas de educação.

Entretanto, a proficiência média de uma escola, que é o resultado da média das proficiências dos estudantes, pode ser insuficiente para capturar as nuances de aprendizagem, principalmente quando os níveis que classificam os alunos como "proficientes"ou "avançados" estão situados em faixas de pontuação relativamente baixas. Isso pode criar uma falsa impressão de progresso educacional, não refletindo o desenvolvimento de habilidades complexas e profundas; corre-se o risco de criar uma percepção de sucesso que não condiz plenamente com a aquisição de habilidades essenciais. Além disso, ao suavizar a diferença entre níveis, especialmente nas faixas inferiores, é possível que lacunas significativas de aprendizado passem despercebidas, prejudicando a implementação de políticas que realmente atendam às necessidades dos alunos. Dessa forma, a interpretação da proficiência precisa ser criteriosa para evitar a distorção de resultados.

Nesse estudo, a proficiência e análise de desempenho levará em consideração as notas contínuas da escolas e não suas faixas ou níveis de classificação a fim de mitigar esse viés.

2.1.3 Composição dos questionários e bases do SAEB

O conjunto de dados da SAEB (os microdados) é composto de bases de identificação e bases advindas de questionários complementares às avaliações e suas respectivas notas. Os questionários são aplicados com o objetivo de coletar dados que possam contextualizar os resultados dos testes cognitivos do SAEB e proporcionar informações e indicadores que permitam avaliar as diferentes dimensões da qualidade da educação, como Equidade, Direitos Humanos e Cidadania, Ensino-Aprendizagem, Investimento, Atendi-

mento Escolar, Gestão e Profissionais Docentes, conforme a matriz de referência do SAEB. Cada base tem suas variáveis e formato de coleta/tratamento descritas no "dicionário de dados", disponível nos Documentos de Resultados do SAEB[10] e no "Leia-me Microdados SAEB 2019"[8].

A descrição das principais bases do SAEB, e a composição de cada base ou tipo de questionário é tal:

- **Questionários de Estudantes:** Há três questionários distintos para estudantes: um para o 5º ano do Ensino Fundamental, outro para o 9º ano, e um terceiro para as 3ª/4ª séries do Ensino Médio. Embora semelhantes, o questionário do 9º ano tem um item adicional, e o das 3ª/4ª séries tem dois itens extras em relação ao 5º ano. Estes questionários coletam dados sobre o contexto socioeconômico, perfil familiar e trajetória escolar dos alunos. A base de dados inclui informações detalhadas sobre o ambiente doméstico e socioeconômico, além de percepções sobre a trajetória educacional dos estudantes.

- **Questionário do Diretor:** Preenchido pelo diretor(a) em formato eletrônico, o questionário do diretor avalia áreas como Gestão (planejamento e participação), Investimento (arrecadação de recursos), Profissionais da Educação (formação, condições de trabalho) e Equidade (inclusão) na escola em que possui o cargo. São coletadas respostas a 210 itens, muitos com regras de dependência ou seja, que dependem da resposta anterior ou se desdobram a partir de uma única questão. Essa base de dados inclui informações sobre a gestão escolar, problemas enfrentados, e programas de apoio à aprendizagem, essenciais para compreender como a gestão afeta o ambiente educacional.

- **Questionário do Professor:** Professores do Ensino Fundamental e Médio preencheram o questionário em papel, enquanto alguns professores da Educação Infantil participaram de um estudo piloto em formato eletrônico. O questionário coleta dados sobre a formação, condições de trabalho, práticas pedagógicas e equidade entre os professores. Essas informações ajudam a analisar como as condições de trabalho e a experiência docente impactam o desempenho dos alunos.

- **Questionário do Secretário de Educação:** Voltado para gestores municipais, o questionário abrange tópicos como Gestão, Investimento, Profissionais da Educação,

Equidade e Ensino-Aprendizagem. A aplicação foi feita eletronicamente com apoio de órgãos como a Undime e Confederação Nacional dos Municípios. Devido à baixa adesão de secretários estaduais e ao caráter experimental dos questionários da Educação Infantil, esses dados não foram incluídos nos microdados de 2019.

- **Dados da Turma e da Escola:** Essas bases incluem informações detalhadas sobre a série em que os alunos estão, a dependência administrativa da escola, sua localização e infraestrutura. Esses dados são fundamentais para entender as condições que afetam o desempenho dos estudantes. Também há nessa base as nota médias de cada escola nas avaliações.

- **Dados da Proficiência e Prova/Item** Os dados de proficiência medem o desempenho dos alunos em Língua Portuguesa e Matemática, utilizando a escala do SAEB. Esses dados são essenciais para análises comparativas e de tendências. A base de dados de itens registra as respostas dos alunos e os parâmetros da Teoria de Resposta ao Item (TRI) usada para calibrar os testes.

2.2 Análise exploratória de dados - AED

Segundo Rumsey (2010) [11], estatísticas descritivas são números que resumem algumas características de um conjunto de dados. Elas fornecem informações de fácil compreensão que ajudam os pesquisadores a obter uma ideia geral sobre os dados, permitindo-lhes realizar análises formais e direcionadas. É por meio de estatísticas descritivas e visualizações que a Análise Exploratória de Dados (AED), processo inicial de uma análise de dados, visa melhor entendê-los, ajuda identificar padrões, detectar anomalias e preparar os dados para análises mais profundas. Para Tukey (1977) [12], a AED é comparada ao trabalho de um detetive. Assim como um investigador precisa de ferramentas e compreensão para encontrar pistas, o analista de dados deve ter técnicas adequadas para explorar os dados, como ferramentas gráficas, numéricas e de contagem. Embora a AED seja apenas o primeiro passo na análise de dados, ela é crucial para encontrar padrões e sinais antes de aplicar métodos confirmatórios mais rigorosos.

Nesse estudo, as principais ferramentas de análise exploratória utilizadas incluem a preparação e o tratamento dos dados, além de estatísticas descritivas e métodos gráficos, os quais serão melhor definidos a seguir, conforme Morettin & Singer (2021)[13]:

- **Preparação e tratamento de dados:** é composto pela limpeza de dados, verificação de consistência e remoção de erros. Quando necessário, realiza-se também a transformação de variáveis (como normalização ou criação de variáveis *dummy*) e a detecção de *outliers* (identificação de valores extremos);

- **Uso de estatística descritiva:** Envolve a definição dos tipos de dados, medidas de posição e dispersão, gráficos e correlação. Os dados podem ser **quantitativos**, representando quantidades que podem ser discretas ou contínuas, e **qualitativos**, representando categorias nominais ou ordinais. Para variáveis contínuas, as **medidas de posição** incluem a média, que é o valor central sensível a outliers, a mediana, que é o valor central menos influenciado por outliers, e a moda, que é o valor mais frequente. As **medidas de dispersão** incluem o desvio padrão, que indica o grau de variação dos dados; a variância, que é a média dos quadrados das diferenças em relação à média; e a amplitude, que é a diferença entre o valor máximo e mínimo. Para variáveis categóricas, a análise foca na distribuição de respostas com gráficos de barra. Por fim, a **correlação** mede a relação linear entre duas variáveis contínuas, variando de -1 (correlação negativa) a +1 (correlação positiva), sem implicar causalidade. Ao utilizar o coeficiente de correlação de Pearson para construir uma matriz de correlação, um valor de -1 indica correlação negativa perfeita, enquanto 1 indica correlação positiva perfeita. Métodos gráficos como boxplots, histogramas e gráficos de dispersão também são essenciais para resumir e representar visualmente os dados, facilitando a identificação de padrões e tendências.

- **Análise de Dados:** Pode ser realizada através da análise univariada, bivariada ou multivariada. A **análise univariada** examina uma única variável utilizando gráficos, como histogramas, e medidas descritivas para explorar a distribuição dos dados. Este processo envolve a coleta, sumarização e interpretação de dados de uma única variável, com a primeira etapa sendo a exploração dos dados brutos. Já a **análise bivariada** foca na relação entre duas variáveis e é dividida em três tipos principais: para a análise qualitativa x quantitativa, utilizam-se boxplots para comparar a distribuição de uma variável quantitativa entre diferentes categorias; para a análise quantitativa x quantitativa, usam-se gráficos de dispersão para examinar as relações entre duas variáveis numéricas, observando a direção, forma e força da relação; e para a análise qualitativa x qualitativa, são construídas tabelas

de contingência para entender a relação entre duas variáveis categóricas.

2.3 Árvores de regressão

2.3.1 Histórico e definições gerais

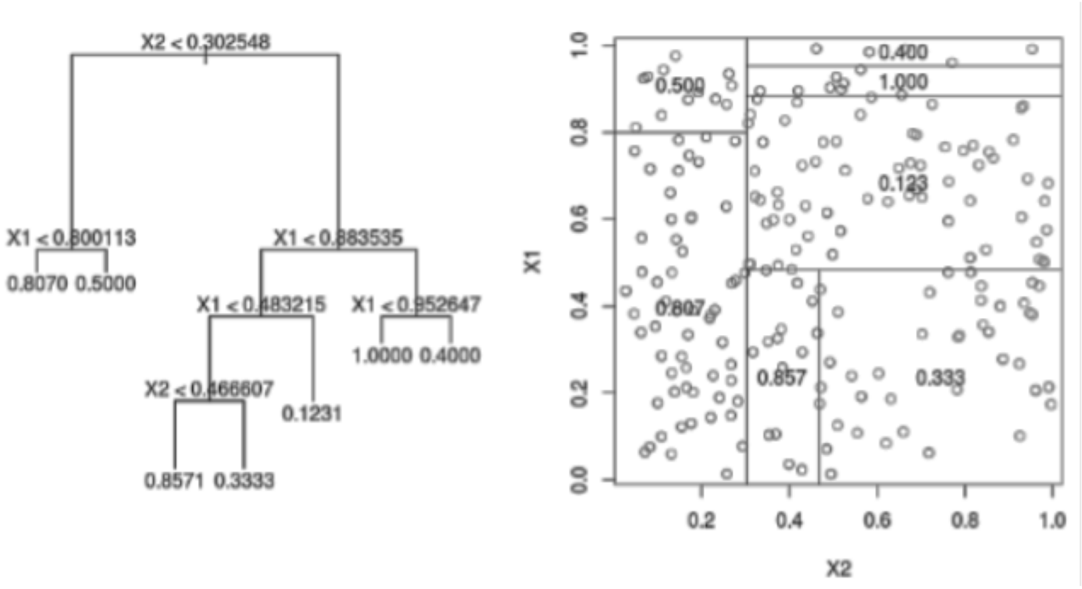
As árvores de decisão são técnicas de modelagem preditiva amplamente usadas nas áreas de estatística, ciência da computação e aprendizado de máquina. Seguindo a definição dada por James et al. (2013) [14], as árvores de decisão buscam estratificar ou segmentar o espaço dos preditores em regiões simples seguindo regras de divisão e que podem ser representadas em uma estrutura de árvore.

Árvores de classificação e de regressão são vertentes das árvores de decisão. As primeiras, visam prever o estado para variáveis alvo categóricas. Já as árvores de regressão, método do estudo aqui desenvolvido, visam prever valores de variáveis-alvo contínuas.

Uma árvore de regressão pode ser vista como um processo de tomada de decisão estruturado: começa-se com o "nó raiz"que abrange toda a população ou amostra de um conjunto de dados e, com base em critérios específicos, esses dados são particionados repetidamente em grupos cada vez mais homogêneos. Cada partição, "nó"ou "divisão", representa uma condição aplicada a uma variável de entrada que separa os dados em diferentes "ramos". No final dos ramos, encontram-se os "nós terminais"ou "folhas", onde a média dos valores da variável-alvo é utilizada para fazer previsões. Esse esquema ilustrativo encontra-se na Figura 2.

Nessa figura é possível verificar as partições (à esquerda) e estrutura gráfica (à direita) da árvore de decisão para um modelo fantasia de regressão com duas variáveis preditoras envolvidas X1 e X2 contínuas. Em cada nó intermediário, um caso vai para o nó filho à esquerda se, e somente se, a condição/regra for satisfeita. Assim, a árvore estratifica a variável de saída em sete regiões de previsão (cada quadrado do esquema à direita) de acordo com as condições dos preditores. Por fim, a previsão de valores contínuos é indicada abaixo de cada nó terminal.

Figura 2: Esquema ilustrativo da uma árvore de regressão e suas partições



Fonte: Data Camp - Decision Trees in Machine Learning Using R[15]

As árvores de regressão, desde sua criação, têm evoluído substancialmente e ampliado sua aplicação e funcionalidade (Wei-Yin Loh, 2014)[16]. O primeiro algoritmo de árvore de regressão foi publicado há mais de cinquenta anos, em 1963, por Morgan e Sonquist, com o método AID (Automatic Interaction Detection). Esse método inicial visava dividir os dados em subconjuntos homogêneos, minimizar a variação dentro de cada nó e gerar previsões quantitativas (para a variável de saída numérica) para cada grupo. Entretanto, o método enfrentava desafios como o excesso de ajuste aos dados e a dificuldade em lidar com variáveis altamente correlacionadas, o que, muitas vezes, mascarava a relevância de certos preditores.

Ao longo dos anos 1970, avanços como o algoritmo THAID (Theta Automatic Interaction Detection) expandiram o uso das árvores para a classificação (em que a variável de saída era categórica), introduzindo métricas de impureza alternativas, como entropia e índice de Gini, que tornaram as divisões mais eficazes na organização dos dados por categorias. Brevemente, Entropia e Índice de Gini são métricas usadas em árvores de decisão para avaliar a impureza dos nós, onde entropia mede a desordem ou incerteza nas classes, e Índice de Gini quantifica a probabilidade de uma classificação incorreta ao

escolher aleatoriamente entre as classes. Esses conceitos não serão aprofundados, pois relacionam-se com árvores de classificação e não de regressão.

Porém, esses métodos encontravam também barreiras computacionais e, apenas em 1984, com o desenvolvimento do método CART (Classification and Regression Trees) por Breiman et al., que o interesse pelas árvores de decisão e regressão realmente se expandiu. O algoritmo CART trouxe melhorias fundamentais, como a técnica de poda para reduzir o sobreajuste e pontuações de importância para as variáveis.

Para Wei-Yin Loh (2014)[16], o CART consolidou o uso das árvores de decisão e regressão. O algoritmo CART e seus sucessores aprimoraram tanto a robustez quanto a aplicabilidade dessas técnicas e permitiram que as árvores de regressão se tornassem modelos mais flexíveis, capazes de ajustar quase todo tipo de modelo estatístico tradicional. Com o tempo, os algoritmos de árvore de regressão se tornaram mais sofisticados, capazes de se adaptar a uma ampla gama de problemas. Hoje, as árvores de regressão são fundamentais para a modelagem preditiva e continuam a evoluir, incorporando abordagens de aprendizado de máquina e estatísticas avançadas, o que amplia seu alcance e suas aplicações.

O método CART será utilizado nesse estudo, está disponível em *software* comercial e pode ser implementado com a biblioteca *rpart* no sistema R.

2.3.2 Algoritmo

De forma sucinta, **o pseudocódigo para a construção de uma árvore de regressão por CART**, começando pelo nó raiz, é dado como (James et al.,2013) [14]:

1. Divida o espaço dos preditores, isto é, o conjunto de valores possíveis para as variáveis $X_1, X_2, \ldots, X_p$, em duas regiões distintas e não sobrepostas R_1, R_2 criando dois nós filhos, de forma a minimizar o critério de divisão (soma dos desvios quadráticos nos dois nós filhos).

2. Para cada observação que pertença a uma região R_j, a previsão será a mesma, e corresponderá à média dos valores de resposta das observações em R_j. Ou seja, o valor predito em cada nó terminal é a média amostral do nó que resulta em uma estimativa constante por partes da função de regressão.

3. Se um critério de parada for atingido - como a profundidade máxima da árvore,

o número mínimo de observações em um nó terminal alcançados ou se a redução de impureza for menor que uma fração predeterminada da impureza no nó raiz - encerre. Caso contrário, aplique o passo 1 recursivamente a cada nó filho.

Ao construir as regiões R_1 e R_2 o objetivo é encontrar uma divisão no espaço dos preditores que minimize a **soma dos quadrados dos resíduos - *residual sum of squares* - (RSS)**, dada pela Equação 1. Em termos práticos, isso significa criar duas regiões o mais homogêneas possível dentro de seus limites para a variável de saída.

$$RSS = \sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \tag{1}$$

Em que y_i é o valor da variável de saída para a i-ésima observação e $\hat{y}_{R_j}$ é a média de resposta para as observações dentro da j-ésima região. Ao considerar todas as partições possíveis do espaço de características em apenas duas regiões, como uma linha única de divisão, adota-se uma abordagem conhecida como **divisão binária recursiva**.

A abordagem de divisão binária recursiva é de cima para baixo - começa no topo da árvore (onde todas as observações pertencem a uma única região) - e, em seguida, divide sucessivamente o espaço dos preditores; cada divisão é indicada por dois novos ramos mais abaixo na árvore. E é uma abordagem com método de seleção localmente ótimo: a cada etapa do processo de construção da árvore, a melhor divisão é feita naquele momento específico, em vez de olhar adiante e escolher uma divisão que levará a uma melhor árvore em uma etapa futura.

Para realizar a divisão binária recursiva, primeiro é selecionado o preditor ou variável X_j e o ponto de corte s de modo a dividir o espaço dos preditores nas regiões $R_1(j, s) = \{X|X_j < s\}$ e $R_2(j, s) = \{X|X_j \geq s\}$ - isto é, em que X_j assume um valor menor que s e maior ou igual a s - leve à maior redução possível no RSS. Nesse processo, consideram-se todos os preditores $X_1, \ldots, X_p$ e todos os valores possíveis do ponto de corte s para cada um dos preditores, mas só são escolhidos o preditor e o ponto de corte que resultem na divisão com o menor RSS.

Aqui, vale ressaltar que o ponto de corte s é um número para variáveis contínuas e uma forma de agrupamento de categorias para variáveis categóricas. Por exemplo, caso uma variável categórica possua 3 categorias, a região 1 será relacionada a uma categoria única e a região 2 será relacionada a uma concatenação das demais categorias, ou seja,

limita-se o agrupamento a dois grupos ou duas partições com elementos excludentes.

Para pontos de corte procurados em variáveis contínuas, ao buscar a minimização do RSS, normalmente são testados somente os pontos médios entre os valores ordenados assumidos pela variável e não todos os valores possíveis para essa variável. Para cada ponto médio calculado, simula-se uma divisão do conjunto de dados em duas partes: uma parte com a variável assumindo valores menores que o ponto médio e outra parte para valores maiores ou iguais a esse ponto. Essa abordagem é mais eficiente, pois reduz o número de possíveis divisões a serem avaliadas, focando apenas nos pontos que representam mudanças significativas nas previsões. Já para categóricos, todas as combinações binárias das categorias - organizadas como foi descrito anteriormente - são testadas para obter uma divisão do espaço que minimize o RSS.

Adaptando-se a equação do RSS para as duas regiões construídas com o corte s, tem-se:

$$RSS = \sum_{i:x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 \tag{2}$$

Em que $\hat{y}_{R_1}$ é a média de resposta para as observações em $R_1(j, s)$, e $\hat{y}_{R_2}$ é a média de resposta para as observações em $R_2(j, s)$. Quão menor sejam as diferenças entre as observações dentro de uma região, menor será o RSS. Quão maior seja o número de características/variáveis p envolvidas no problema, o tempo para encontrar os valores de j e s que minimizem o erro aumentará (James et al.,2013) [14].

Em seguida, repete-se o processo, e procura-se o melhor preditor e o melhor ponto de corte para dividir ainda mais os dados, de modo a minimizar o RSS dentro de cada uma das regiões resultantes. No entanto, desta vez, em vez de dividir todo o espaço dos preditores, dividimos uma das duas regiões identificadas anteriormente. Agora temos três regiões. Novamente, procuramos dividir uma dessas três regiões ainda mais, para minimizar o RSS. O processo continua até que um critério de parada seja alcançado.

A cada repetição do processo, todas as combinações de variáveis e pontos de corte ainda não utilizados na árvore podem ser considerados para uma nova divisão. Assim, caso alguma variável já tenha sido utilizada como preditor de um nó, mas houver disponibilidade de algum ponto de corte ainda não utilizado para a divisão da árvore, não há restrição e essa variável poderá aparecer novamente na árvore em algum momento posterior.

Por fim, uma vez que as regiões R_1, R_2 forem criadas, a previsão da resposta para uma dada observação de teste será dada com a média das observações de resposta na região à qual essa observação pertence.

2.3.3 Poda da árvore por complexidade

Como retomado por James et al.(2013) [14] e proposto por Breiman (1984)[17], o algoritmo inicial tende a superajustar o conjunto de treinamento, resultando em baixo desempenho em novos dados. Isso ocorre porque a árvore resultante pode ser complexa demais caso as regras de parada estejam mal calibradas ou caso não haja uma regra de parada, o que leva ao superajuste. Para contornar isso, recomenda-se a poda da árvore.

Existem outras formas de fazê-la, mas uma estratégia comum será apresentada: deve-se criar uma árvore muito grande T_0 e, em seguida, podá-la para obter uma subárvore que leve à menor taxa de erro no conjunto de teste. A partir da árvore inicial, aplica-se a poda por custo-complexidade - também conhecida como poda pelo elo mais fraco. Em vez de considerar todas as subárvores possíveis, são consideradas uma sequência de árvores indexadas por um parâmetro de ajuste não-negativo α. Um novo **algoritmo proposto para a construção de uma Árvore de Regressão com poda** é dado a seguir:

1. Use a divisão binária recursiva para crescer uma grande árvore nos dados de treinamento, parando apenas quando cada nó terminal tiver menos de um certo número mínimo de observações.

2. Aplique a poda por custo-complexidade à grande árvore para obter uma sequência de melhores subárvores, como função de α.

3. Use validação cruzada para escolher α. Isto é, divida as observações de treinamento em K grupos. Para cada $k = 1, \ldots, K$: Repita os Passos 1 e 2 em todos os grupos, exceto no k-ésimo grupo dos dados de treinamento. Avalie o erro médio de previsão quadrático nos dados do k-ésimo grupo, como função de α. Em seguida, calcule a média dos resultados para cada valor de α e escolha o α que minimiza o erro médio.

4. Retorne a subárvore do Passo 2 que corresponde ao valor escolhido de α.

Para cada valor de α corresponde uma subárvore $T \subset T_0$ tal que o novo RSS considera um custo adicional, o que pode ser visto na Equação 3:

$$RSS_{custo} = \sum_{m=1}^{|T|} \sum_{i:x_i \in R_m} (y_i - \hat{y}_{R_m})^2 + \alpha|T| \tag{3}$$

Espera-se que esse custo seja tão pequeno quanto possível para a árvore escolhida. Aqui, $|T|$ indica o número de nós terminais da árvore T, R_m é o o subconjunto do espaço do preditor correspondente ao m-ésimo nó terminal, e $\hat{y}_{R_m}$ é a resposta prevista associada a R_m, isto é, a média das observações de treinamento em R_m. O parâmetro de ajuste α controla o equilíbrio entre a complexidade da subárvore e seu ajuste aos dados de treinamento. Quando $\alpha = 0$, então a subárvore T será simplesmente igual a T_0. No entanto, à medida que α aumenta, há um custo para se ter uma árvore com muitos nós terminais, e, portanto, ramos são podados da árvore, de modo que obter toda a sequência de subárvores como função de α é simplificada.

Assim, ajustamos α para obter a subárvore ideal, aplicando-a ao conjunto de dados completo. Por fim, vale ressaltar que existem outras formas de podar a árvore, como a definição de uma profundidade máxima para tal ou uma quantidade mínima de observações em um nó terminal. Esses diferentes parâmetros podem ser definidos durante a modelagem da árvore e adicionam valor ao modelo, deixando-o mais personalizado às necessidades da análise quando utilizados em conjunto com a poda por complexidade.

2.3.4 Avaliação de precisão

Na modelagem de regressão, as métricas de avaliação de desempenho possuem um papel fundamental na análise da adequação de um modelo aos dados. Em problemas de regressão, os conceitos de erro e resíduo são centrais para essas métricas. O erro refere-se à diferença entre os valores que o modelo prevê e os valores reais, enquanto o resíduo é a diferença entre os valores observados e os valores previstos dentro do conjunto de treinamento.

Para Willmott (2005) [18], existem diversas métricas usadas para avaliar o desempenho de modelos de aprendizado de máquina, e a escolha da métrica mais adequada depende do tipo de problema e do objetivo do modelo. A seguir são apresentadas algumas das métricas mais utilizadas para avaliar modelos de regressão:

- **Erro Médio Absoluto (Mean Absolute Error - MAE)**: é a média da soma dos erros absolutos entre as previsões e os valores reais. A definição do MAE pode

ser visualizada na Equação 4:

$$MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} \tag{4}$$

Em que:

- n: Tamanho da amostra.

- $\hat{y}_i$: Valor previsto para a i-ésima observação.

- y_i: Valor correspondente à i-ésima observação.

- $|\hat{y}_i - y_i|$: Erro absoluto, a diferença entre o valor previsto e o verdadeiro em valor absoluto.

- **Erro Quadrático Médio (Mean Squared Error - MSE)**: é a média da soma dos erros quadraticos entre as previsões e os valores reais. A elevação ao quadrado tem o efeito de penalizar mais erros maiores do que menores. Isso significa que o MSE tende a dar mais peso aos erros grandes. Uma desvantagem é que ele pode ser difícil de interpretar diretamente, pois está em unidades quadradas. A definição do MSE pode ser visualizada na Equação 5:

$$MSE = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n} \tag{5}$$

Em que:

- n: Tamanho da amostra.

- $\hat{y}_i$: Valor previsto para a i-ésima observação.

- y_i: Valor correspondente à i-ésima observação.

- **Raiz do Erro Quadrático Médio (Root Mean Squared Error - RMSE)**: é a raiz quadrada do MSE, indica o erro típico, dando mais peso aos desvios maiores. É comum usar a raiz quadrada do MSE, o RMSE, que tem as mesmas unidades que a variável de destino. A definição do RMSE pode ser visualizada na Equação 6:

$$RMSE = \sqrt{MSE} \tag{6}$$

- **Coeficiente de determinação ($\mathbf{R^2}$)**: determina o quanto da variação dos dados foi explicada pelo modelo. Seu valor varia entre 0 e 1, sendo que quanto mais próximo de 1, melhor o ajuste do modelo aos dados observados. Esse coeficiente expressa quão bem o modelo linear explica a variação dos dados em comparação com o modelo baseado apenas na média dos valores observados. Pode ser calculado através das Equações 7, 8 e 9:

  - Soma Total dos Quadrados (SQtot): Representa a soma das diferenças entre cada valor observado y_i e a média dos valores $\bar{y}$:

$$SQ_{\text{tot}} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{7}$$

  Onde: n é o número de observações, y_i é o valor observado e $\bar{y}$ é a média dos valores observados.

  - Soma dos Quadrados dos Resíduos (SQres): Mede a soma das diferenças entre os valores observados y_i e os valores previstos $\hat{y}_i$ pelo modelo:

$$SQ_{\text{res}} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{8}$$

  - R^2 (Coeficiente de Determinação): É dado como complemento da fração dos resíduos em relação ao total:

$$R^2 = 1 - \frac{SQ_{\text{res}}}{SQ_{\text{tot}}} \tag{9}$$

Assim, entende-se que um método de regressão é bom quando os valores de MAE e RMSE são baixos, ou seja, para dois modelos com a mesma base e variáveis alvo, aquele com menor valor de MAE e RSS será o mais adequado para o problema. E para um modelo mais preditivo, mais próximo de 1 o R^2 deve ser.

2.3.5 Justificativas do uso das árvores de regressão

O método de Árvores de Regressão surge como uma abordagem factível devido à sua interpretabilidade e capacidade de lidar com variáveis categóricas e contínuas de

maneira eficiente. Breiman (1984)[17] argumenta que essas árvores oferecem uma alternativa interessante para analisar problemas e podem fornecer pistas sobre a estrutura dos dados que não são aparentes em uma análise de regressão linear. Para esse estudo de caso, com a quantidade de dados tratados - cerca de 37.871 linhas/registros e 200 colunas/variáveis que serão definidas em capítulos posteriores - é provável que o uso de um modelo linear indicasse significância em todas as variáveis preditoras x_1 até x_{200}. Basicamente, em função do tamanho da amostra, não seria razoável apresentar um modelo com tantos componentes e sem a elucidação das variáveis mais relevantes para o desempenho educacional.

Entende-se também que árvores de regressão oferecem uma vantagem considerável em relação a outros métodos não-lineares, como Random Forest ou Bagging, que são comumente vistos como "caixas-pretas"por sua complexidade e dificuldade de interpretação. Assim, as árvores de decisão podem ser facilmente visualizadas e compreendidas, o que ajuda a explicar as decisões tomadas pelo modelo.

Além disso, algumas especifidades desse método para análise de dados são:

- As árvores de decisão realizam naturalmente a seleção de características, escolhendo as características mais informativas para dividir os dados. A importância de uma característica pode ser determinada com base na ordem em que ela aparece na árvore e com que frequência é usada para divisão, apresenta uma espécie de hierarquia;
- É fácil entender quais variáveis são importantes para fazer a previsão - os nós internos (divisões) são aquelas variáveis que mais reduziram o erro.

No entanto, apesar de sua simplicidade e interpretabilidade, as Árvores de Decisão têm algumas limitações. Para Breiman (1984)[17], os modelos de árvores de decisão frequentemente não são tão precisos quanto outros métodos de aprendizado de máquina não lineares. Devido à sua simplicidade de sua construção, as árvores de decisão nem sempre produzem os modelos mais precisos. Todavia, essa questão é compensada pela sua estrutura hierárquica - semelhante à tomada de decisões humanas. Esse ganho de interpretabilidade é crucial em situações como a tratada, em que entender o processo decisório é tão importante quanto a própria previsão.

As árvores de regressão também podem ser sensíveis a variações nos dados ou podem se tornar excessivamente complexas quando crescem demais, o que leva ao *over-*

fitting (ajuste excessivo) do modelo ao conjunto de dados e resulta em uma generalização deficiente de dados não vistos. No entanto, técnicas como a poda podem ajudar a mitigar esse problema.

Assim, considerando os *trade-offs* presentes na escolha do modelo e o contexto deste trabalho, a decisão pelo uso das Árvores de Regressão justifica-se pela necessidade de uma ferramenta interpretável para analisar o vasto conjunto de dados do SAEB e identificar variáveis relacionadas à infraestrutura escolar, gestão e demais características que influenciam o desempenho escolar em preferência ao ganho de poder preditivo de outros modelos.

3 MÉTODOS E PREMISSAS

Esse capítulo descreve como foram implementadas as técnicas da revisão bibliográfica no presente estudo de caso e bases contempladas. Inicia-se a descrição da metodologia pela explicação da fase de obtenção dos dados e os critérios de tratamento e filtragem das informações para a construção de uma base consolidada. Também há a explicação sobre a exploração prévia das informações da base através da estatística descritiva dos dados, e posterior formulação do problema de modelagem, suas premissas e considerações gerais. A metodologia empregada pode ser sintetizada pela Figura 3 - Fluxograma da Metodologia:

Figura 3: Fluxograma da Metodologia

Fonte: Elaboração própria

3.1 Critérios de escolha e forma de obtenção dos dados

A escolha dos dados é uma etapa fundamental em estudos que envolvem grandes volumes de informações, pois a confiabilidade da base é essencial antes de modelar ou compreender qualquer fenômeno. Nesse sentido, é crucial considerar os cinco "Vs"do *Big Data*: volume, que se refere à quantidade massiva de dados; velocidade, relacionada à rapidez com que os dados são gerados e processados; variedade, que trata da diversidade de formatos e fontes; veracidade, que avalia a precisão e confiabilidade dos dados; e valor, que corresponde ao potencial dos dados para gerar insights. Esses elementos são indispensáveis para garantir a qualidade e relevância dos dados utilizados na análise, como destacado por Bourany (2018)[19], ao abordar os desafios que a era digital impõe na coleta e no uso de dados confiáveis e precisos.

O conjunto de bases do SAEB foi escolhido como objeto de estudo e análise devido à confiabilidade e robustez dos dados, sendo que os cinco "Vs"do *Big Data* foram validados

no contexto deste estudo. O volume de informações disponíveis no SAEB é expressivo, com milhares de registros de escolas e alunos com abrangência nacional. A velocidade com que os dados são gerados e atualizados, devido à periodicidade bienal das avaliações, asseguram que as informações sejam recentes e aplicáveis. A variedade é evidente pela ampla gama de variáveis, que inclui desempenho acadêmico, informações sobre infraestrutura e gestão escolar. A veracidade dos dados foi assegurada pela confiabilidade da fonte, uma vez que o SAEB é uma das principais bases de dados educacionais do Brasil, gerida por um órgão oficial. Por fim, o valor desses dados está na possibilidade de gerar análises multifacetadas e relevantes. Dessa forma, com a quantidade de dados e variáveis disponíveis, é possível utilizar técnicas de regressão e análises estatísticas com maior precisão.

Para esse caso, os dados do SAEB de 2019 foram utilizados. Tendo em vista que as avaliações são realizadas a cada 2 anos, a escolha pelo conjunto de 2019 deu-se a fim de evitar vieses do cenário escolar brasileiro durante a pandemia (o que seria encontrado nas bases de 2021[2], também já disponíveis no site do INEP, mas não utilizadas nesse trabalho). Deste conjunto, foram escolhidas a Base Escolas e a Base Diretor - as quais se relacionam através do ID_ESCOLA e que possuem mais de 40 mil registros únicos de cada escola, com mais de 200 variáveis/colunas com informações sobre elas e as notas de cada uma.

Após a obtenção das bases, foi necessário consolidá-las e iniciar o tratamento inicial dos dados. Para cada escola, o SAEB apresenta as notas de avaliações de língua portuguesa e matemática para as turmas de 5º e 9º anos do ensino fundamental e para o 3º ano do Ensino Médio. Para garantir um estudo focado e aprofundado sobre o desempenho das escolas, optou-se por modelar o problema concentrando-se em uma única variável dependente. Assim, foram selecionados exclusivamente os dados referentes às notas do 5º ano, pois é no estágio inicial da vida escolar que eventuais problemas educacionais começam a se manifestar de forma mais clara. O objetivo dessa filtragem foi evitar uma abordagem dispersa, sem a devida precisão.

[2]Em entrevista à CNN[20], o diretor-executivo do Todos Pela Educação, Olavo Nogueira Filho, ressaltou a importância de cautela ao analisar os resultados de 2021, mencionando que a aplicação das provas foi fortemente impactada pela pandemia e gerou distorções atípicas. Segundo o especialista, esses dados "demandariam um olhar muito cuidadoso", especialmente em comparações entre redes, devido às questões atípicas influenciadas pelo contexto.

3.2 Análise exploratória e preparação dos dados

A fase da análise exploratória de dados (AED), inicia-se com a aplicação de estatística descritiva das informações, sendo necessário identificar o tipo de variáveis presentes na base de dados: distinguir quais variáveis são categóricas, listando a quantidade de níveis, e quais são numéricas. Esse processo envolve entender o conteúdo da base, examinar as perguntas e respostas, e calcular médias e distribuições de cada variável quando possível.

Nesse estudo, a análise de dados foi conduzida principalmente por meio da análise univariada, sem utilizar análise bivariada para examinar como as variáveis independentes se conectam à variável contínua dependente (devido à alta quantidade de variáveis). O foco foi explorar a distribuição das variáveis da base em seus próprios níveis e compreender o que é comum à maioria das escolas ou extraordinário a elas.

Já para a preparação dos dados, com a intenção de definir quais informações entrariam no modelo, foram aplicadas as seguintes estratégias:

- **Transformação de variáveis categóricas politômicas em dicotômicas (quando possível)**, criando apenas dois cenários/níveis para a maioria das variáveis, por exemplo, uma escala de concordância dividida em 4 níveis foi transformada em "Concordo"e "Discordo". Essa medida ajuda a simplificar a estrutura da árvore, reduzir ruído, revelar padrões mais claros e melhorar a eficiência computacional.

- **Remoção de variáveis que apresentavam dependência entre si**, pois respostas advindas de perguntas dependentes possuem um número reduzido de respondentes, assim, dados faltantes (*missing data*) atrapalhariam ao reduzir a precisão dos modelos e dificultar análises.

- Para identificar as relações entre as variáveis contínuas independentes com a variável resposta, foram **utilizadas matrizes de correlação**, as quais ajudam a identificar e remover recursos altamente correlacionados, reduzir o risco de sobreajuste e diminuir o tempo de execução de modelos. É ideal que as variáveis tenham uma relação clara com a variável resposta ou entre si, para que possam ser destacadas do modelo ou mantidas.

3.3 Ciclos de modelagem e avaliação

A penúltima etapa do processo é a avaliação e seleção final do modelo, onde ciclos de modelagem são realizados até que se obtenham resultados satisfatórios e adequados aos objetivos do estudo.

Neste estudo, a modelagem foi realizada no ambiente RStudio, utilizou-se a biblioteca *rpart* para construir modelos de árvores de regressão em R. Uma das vantagens de utilizar o *rpart* para árvores de regressão em R é que não há necessidade de criar variáveis *dummy* ao lidar com variáveis categóricas, ou seja, não é necessário utilizar *One-Hot Encodind* para codificação dos dados de entrada - procedimento necessário ao utilizar bibliotecas do Python e que adiciona complexidade ao processo. O *rpart* lida nativamente com essas variáveis, tratando-as como grupos distintos e considerando todas as possíveis divisões desses grupos ao construir a árvore. Assim, as variáveis categóricas da base podem ser incluídas diretamente no modelo sem a necessidade de transformação.

Dessa forma, após a filtragem da base para utilização somente das variáveis selecionadas, como não há a necessidade de transformações dos dados de entrada, é realizada a separação dos dados em conjuntos de treinamento e teste. Essa separação permite que o modelo seja treinado em um subconjunto dos dados e, em seguida, avaliado em outro subconjunto independente, para garantir que o modelo possa desempenhar bem para novos dados. A fração de dados utilizada foi de 80% para treino e 20% para teste (dado que a modelagem visa focar na melhor descrição dos dados em vez de focar na previsão com alta precisão de uma nota para uma nova escola). Com a fração de dados para teste, calcula-se o erro médio com a diferença entre as notas reais das escolas e os valores previstos pelo modelo. Para garantir a reprodutibilidade dos resultados, foi utilizado uma semente *set.seed* que dividiu as amostras.

Com os dados de treinamento preparados, o próximo passo foi ajustar o modelo utilizando o *rpart*. Esse ajuste consiste em aplicar o algoritmo de árvore de regressão aos dados para identificar a melhor estrutura de árvore possível, podá-la em profundidade, utilizar a técnica de validação cruzada (divide-se o conjunto de dados em múltiplos subconjuntos, ajusta-se e testa-se o modelo em cada combinação de subconjuntos) e ajustar o seu tamanho ao custo de complexidade, entre outros. Após o ajuste, o modelo é testado com o conjunto de dados de teste, que não foi utilizado durante o processo de treinamento.

A avaliação do modelo foi realizada com base em métricas como o Erro Médio

Absoluto (MAE), Raiz do Erro Quadrático Médio (RMSE) e o coeficiente de determinação (R^2). O MAE foi utilizado para medir a média das diferenças absolutas entre os valores reais e previstos, enquanto o RMSE indicou o erro típico, dando mais peso aos desvios maiores. Além disso, o R^2 foi calculado para determinar o quanto da variação dos dados foi explicada pelo modelo, sendo essas métricas essenciais para avaliar a precisão e a eficácia do modelo em prever novos resultados.

Após os ciclos de modelagem e avaliação, a última etapa do processo é então a interpretação dos resultados das árvores. Tendo em vista que as árvores são ferramentas bastante visuais, é possível compreender as características das escolas com melhores e piores desempenhos e descrever os efeitos identificados.

4 RESULTADOS PRÉ-MODELAGEM

Este capítulo apresenta os resultados obtidos nas etapas pré-modelagem do estudo e parte da estruturação da base e sua análise exploratória para contextualização da situação educacional indo até a seleção das variáveis para o modelo. Cabe ressaltar que as tabelas e relações a seguir são de elaboração própria, nas quais foram utilizados os dados advindos do SAEB como fonte.

4.1 Estruturação da base

Seguindo a metodologia proposta, a primeira etapa de estruturação de uma base de trabalho, foi a junção das duas bases do SAEB mencionadas: a base ESCOLAS e a base DIRETOR. Na primeira, foram obtidas as informações geográficas das escolas, como sua região, nível socioeconômico, localização urbana ou rural, além das notas específicas para as avaliações, como a média de língua português e de matemática para os alunos do quinto ano do ensino fundamental. Aqui foram filtrados somente os registros que possuíam dados de nota para português e matemática no quinto ano, totalizando 42.082 escolas.

Já na base DIRETOR, as informações retiradas foram referentes aos questionários de cada escola, contendo as colunas foram referentes às respostas das 251 questões sobre a rotina da direção, a infraestrutura da escola, os programas pedagógicos entre outros. Aqui, foram filtradas somente as escolas que possuíam valor de coluna igual a 1 na IN_PREENCHIMENTO_QUESTIONARIO, indicando que o questionário havia sido parcialmente ou totalmente preenchido e totalizando 37.871 escolas válidas para análise - cerca de 90% das 42.082 escolas filtradas em notas.

Para juntar as informações dessas duas bases, foi utilizada a coluna ID_ESCOLA, presente em ambas como um identificador de registro de escola. Com essa nova base de dados consolidada, foi possível iniciar a análise descritiva prévia das informações.

4.2 Análise exploratória univariada

Para a análise exploratória com uso de estatística descritiva da base consolidada, essa base foi dividida em 7 subpartes, assim como é apresentado no dicionário de dados fornecido pelo SAEB. A seguir, serão apresentadas a distribuição dos registros em cada

variável da base - aqui, ainda não é desconsiderada nenhuma das colunas originais e busca-se compreender a composição da base.

4.2.1 Informações demográficas

Para as informações relacionadas à distribuição espacial das escolas em regiões temos a Tabela 1 (Distribuição de escolas por região). Nela, é possível compreender que há uma concentração maior de escolas no nordeste e sudeste, seja em escolas válidas ou não.

Tabela 1: Distribuição de escolas por região

Região	Contagem Escolas Totais	%	Região	Contagem Escolas Válidas	%
Nordeste	14.631	35%	Nordeste	13.567	36%
Sudeste	13.269	32%	Sudeste	12.167	32%
Sul	6.670	16%	Sul	5.499	15%
Norte	4.562	11%	Norte	4.018	11%
Centro-Oeste	2.950	7%	Centro-Oeste	2.620	7%
Total Geral	42.082	100%	Total Geral	37.871	100%

Para a distribuição de escolas por estado, temos a Tabela 2 (Distribuição de escolas por estado), na qual é possível verificar que o Rio de Janeiro é o estado com mais escolas por município e também um dos estados com maior densidade. Com essa tabela é possível compreender que nem todos os estados possuem uma densidade escolar semelhante, alguns parecem carecer de escolas devido a sua extensão territorial ou alta quantidade de municípios.

Tabela 2: Distribuição de escolas por estado

Estado	Contagem Escolas Totais	Municípios	Escolas/Município	Área Estado (km^2)	Escolas/km^2 (10^3)
RJ	2.266	92	25	43.750	52
PA	2.279	144	16	1.245.871	2
AM	912	62	15	1.559.256	1
CE	2.366	184	13	148.894	16
AP	197	16	12	142.471	1
ES	830	78	11	46.074	18
PE	1.921	184	10	98.068	20
AC	229	22	10	164.173	1
MA	2.255	217	10	329.651	7
DF	338	35	10	5.761	59
SP	5.765	645	9	248.219	23
AL	895	102	9	27.831	32
BA	3.608	417	9	564.760	6
SE	627	75	8	21.938	29
RO	427	52	8	237.754	2
MS	592	79	7	903.208	1
RR	101	15	7	223.645	0
PR	2.427	399	6	199.299	12
SC	1.704	295	6	95.731	18
MT	782	142	6	357.142	2
MG	4.408	853	5	586.514	8
RS	2.539	497	5	281.707	9
GO	1.238	246	5	340.243	4
RN	830	167	5	52.810	16
PB	1.081	223	5	56.467	19
PI	1.048	224	5	251.755	4
TO	417	139	3	277.424	2
Total geral	42.082	5.604	-	8.510.418	-

Para a distribuição de escolas por área e localização, temos as Tabelas 3 e também a 4 . Com elas, nota-se que a grande maioria das escolas - cerca de 88% está localizada fora das capitais dos estados, e que somente 25% das escolas localiza-se em área rural, o que mais uma vez poderia indicar uma demanda mal suprida de escolas pela população de áreas afastadas ou descentralizadas.

Tabela 3: Distribuição de escolas por área

Área	Contagem Escolas Totais	%	Contagem Escolas Válidas	%
Capital	5.247	12%	4.481	13%
Interior	36.835	88%	32.990	87%
Total geral	42.082	100%	37.871	100%

Tabela 4: Distribuição de escolas por localização

Localização	Contagem Escolas Totais	%	Contagem Escolas Válidas	%
Rural	10.724	25%	9.281	25%
Urbana	31.358	75%	28.590	75%
Total geral	42.082	100%	37.871	100%

Ainda sobre a distribuição escolar, em relação aos níveis socioeconômicos de cada escola, propostos pelo Indicador de Nível Socioeconômico (Inse)[21] do SAEB 2019 - os quais auxiliam na identificação das desigualdades educacionais e relacionam-se a quantidade de bens materiais nas casas dos alunos frequentadores das escolas e ao nível de escolaridade dos pais dos alunos) - é possível verificar a concentração de escolas nos níveis médios III, IV e V. Já por região, existe uma concentração acentuada de escolas de maior nível socioeconômico nas regiões Sudeste e Sul, também há concentração de escolas de menor nível socioeconômico nas regiões Norte e Nordeste. Essas informações podem ser visualizadas na Tabela 5

Tabela 5: Distribuição do Nível Socioeconômica e Regional de escolas

Nível Socioeconômico	Norte	Nordeste	Centro-Oeste	Sudeste	Sul	Total	% Nível
Nível I	16	12	-	-	-	28	0%
Nível II	1.106	3.181	6	110	1	4.404	10%
Nível III	1.892	7.719	144	789	66	10.610	25%
Nível IV	1.326	3.368	1.430	4.728	958	11.810	28%
Nível V	156	121	1.248	6.444	3.766	11.735	28%
Nível VI	14	5	118	1.145	1.806	3.088	7%
Nível VII	-	1	1	27	50	79	0%
Não Informado	52	224	3	26	23	328	1%

Já em relação à participação dos alunos nas provas do SAEB, a Tabela 6 apresenta

a quantidade de alunos matriculados em cada escola participante e também a quantidade de alunos presentes para a realização da prova. Assim, foi possível estimar o porte médio das escolas de cada estado para o quinto ano do ensino fundamental - em que todos os estados apresentam porte na casa das dezenas de alunos - e também estimar a taxa de participação dos alunos na prova - que esteve maior que 92% para todos os estados, indicando uma alta aderência ao sistema de avaliação.

Tabela 6: Porte e Participação Escolar no SAEB

UF	Matriculados Censo 5EF	Presentes SAEB 5EF	Contagem de Escolas Totais	Porte médio (matriculados/escolas)	Taxa de Participação (presentes/matriculados)
AC	13.111	12.161	229	57	93%
AL	41.963	39.732	895	47	95%
AM	58.023	54.789	912	64	94%
AP	12.123	11.208	197	62	92%
BA	149.111	138.994	3.608	41	93%
CE	101.120	99.712	2.366	43	99%
DF	30.851	28.694	338	91	93%
ES	43.632	40.164	830	53	92%
GO	72.864	68.061	1.238	59	93%
MA	90.845	85.301	2.255	40	94%
MG	221.288	210.804	4.408	50	95%
MS	34.638	32.128	592	59	93%
MT	40.670	38.483	782	52	95%
PA	118.213	110.219	2.279	52	93%
PB	40.019	37.665	1.081	37	94%
PE	95.237	90.516	1.921	50	95%
PI	38.579	36.026	1.048	37	93%
PR	125.920	11.9863	2.427	52	95%
RJ	131.271	121.377	2.266	58	92%
RN	32.969	30.036	830	40	91%
RO	24.897	23.606	427	58	95%
RR	6.587	6.075	101	65	92%
RS	83.965	77.130	2.539	33	92%
SC	75.867	70.977	1.704	45	94%
SE	24.593	22.754	627	39	93%
SP	457.433	429.627	5.765	79	94%
TO	23.707	22.091	417	57	93%
Total	2.189.496	2.058.193	42.082	52	94%

Por fim, para as informações demográficas, tem-se as notas das avaliações para as escolas em língua portuguesa e em matemática. Nas Figuras 4 - Histograma Notas LP e 5 - Histograma Notas MT, é possível verificar que, para as 37.871 escolas válidas, a média em português foi de 206 pontos e em matemática 219 pontos. Ao considerar as notas contínuas em faixas de desempenho propostas pelo próprio SAEB - nas Tabelas 7 e 8 - ainda que grande parte das escolas esteja classificada em níveis de proficiência, para língua portuguesa, 38% delas está abaixo do nível proficiente e para matemática o cenário

é tal que 57% das escolas está abaixo dessa linha. Essa situação demonstra a necessidade de identificar possíveis fatores que afetam diretamente o desempenho das escolas.

Figura 4: Histograma Notas LP

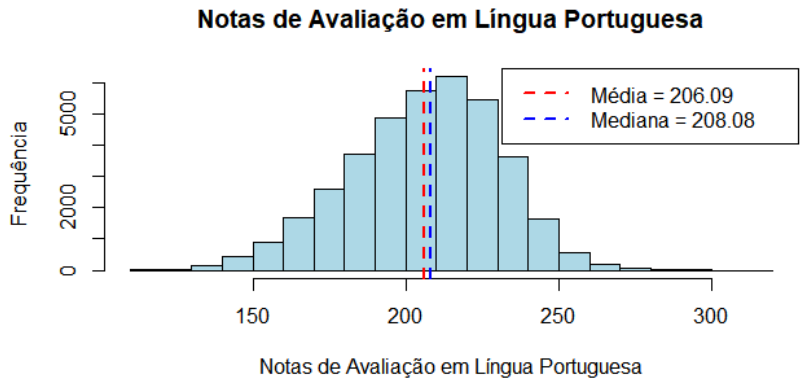


Figura 5: Histograma Notas MT

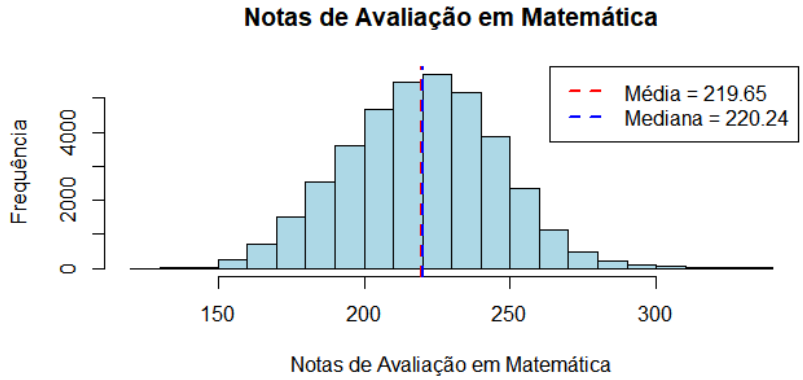


Tabela 7: Distribuição de escolas por faixa de aprendizado - Língua Portuguesa

Faixa de Língua Portuguesa	Contagem de Escolas Total	%
Insuficiente	611	2%
Básico	13.749	36%
Proficiente	22.651	60%
Avançado	860	2%
Total	37.871	100%

Tabela 8: Distribuição de escolas por faixa de aprendizado - Matemática

Faixa de Matemática	Contagem de Escolas Total	%
Insuficiente	1.652	4%
Básico	20.023	53%
Proficiente	15.616	41%
Avançado	580	2%
Total	37.871	100%

4.2.2 Informações do diretor

A partir dessa subseção, inicia-se a descrição das respostas obtidas através do questionário da base DIRETOR. Para cada uma das 251 questões, a coluna TX_RESP_Q001 a TX_RESP_Q251 armazena as respostas. Vale ressaltar que algumas das questões sofrem regras de dependência de questões anteriores, isto é, só podem ter resposta armazenada caso haja resposta específica da questão "mãe", a qual dá sentido às questões dependentes. Nesses casos, a quantidade de respostas válidas, que gira em torno das 37.871, cai de forma drástica para menos de 50% desse valor, o que poderá ser visualizado nas tabelas.

As tabelas dessa e das próximas seções da análise descritiva concentram questões tanto por macro quanto por subtemas, mas que sempre possuem os mesmo tipos de respostas se em uma única tabela. Para questões com resposta aberta contínua, foram criadas tabelas contendo as média e mediana das respostas e informações. Em casos de tabelas contendo muitas questões, pode haver classificação e ordenamento dessas de acordo com a quantidade de respostas de um tipo obtidas a fim de auxiliar na visualização dos dados e informações relevantes do tema.

Iniciando a análise pelas informações sobre o diretor das escolas e sua experiência destaca-se que 45% dos diretores autodeclaram-se brancos e 44% pardos, como pode ser visto na na Tabela 9 (TX_RESP_Q001 - Cor/Raça do Diretor). Além disso, na Tabela 10 (TX_RESP_Q002 a TX_RESP_Q021 - Experiência e rotina do diretor), nota-se que, em média, os diretores possuem 13 anos de experiência como professores antes de assumirem o cargo de diretor, 6 anos de experiência nesse cargo e 4,3 anos como diretores da escola atual pela qual estão respondendo o questionário. O alto desvio padrão dessas respostas indica que as respostas são bastante diversas, ou seja, existem experiências de diretoria

variadas.

Sobre a rotina de uma semana de trabalho, as atividades realizadas estão bem distribuídas, tomando - em média - de 2,9h (atividades relacionadas à segurança da escola) a 6,4h (atendimento aos alunos) das horas de trabalho do diretor.

Tabela 9: TX_RESP_Q001 - Cor/Raça do Diretor

Qual é a sua cor ou raça?	Contagem Respostas	%
Não quero declarar	420	1%
Amarela	390	1%
Branca	17.071	45%
Indígena	183	≈ 0%
Parda	16.543	44%
Preta	3.043	8%
Total	37.650	100%

Tabela 10: TX_RESP_Q002 a TX_RESP_Q021 - Experiência e rotina do diretor

Questão	Descrição		Mínimo	Máximo	Média	Mediana	Desvio padrão	Total
TX_RESP_Q002E3	Por quanto tempo você trabalhou como professor (a) antes de se tornar diretor (a)? Anos		0	35,8	13,5	13	7,8	37.871
TX_RESP_Q004E5	Você possui quanto tempo de experiência como diretor(a) de escola? Anos		0	35,9	6,1	4	5,6	37.871
TX_RESP_Q006E7	Há quanto tempo você é diretor(a) desta escola? Anos		0	35	4,3	3	4,4	37.871
TX_RESP_Q008	Considerando todas as suas atividades profissionais remuneradas, quantas horas você trabalha em uma semana normal?		0	70	42,1	40	9,8	35.984
TX_RESP_Q009	Quantas horas você trabalha em uma semana normal em atividades relacionadas à educação?		0	70	42,2	40	11,2	35.821
TX_RESP_Q010	Na semana normal de trabalho, quantas horas você trabalha para esta escola?		0	70	41,2	40	9,7	36.033
TX_RESP_Q011	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola:	Prestação de contas	0	30	4,3	3	4,2	36.115
TX_RESP_Q012		Reunião com professores	0	20	3,6	3	2,6	36.492
TX_RESP_Q013		Atendimento aos pais ou responsáveis	0	30	5,5	4	4,4	36.119
TX_RESP_Q014		Gerenciamento de conflitos	0	30	5,1	4	4,8	35.922
TX_RESP_Q015		Atendimento aos(às) alunos(as)	0	30	6,4	5	5,2	35.366
TX_RESP_Q016		Atendimento individual aos(às) professores(as)	0	30	4,0	3	3,9	36.159
TX_RESP_Q017		Demandas da Secretaria de Educação	0	30	5,9	4	4,8	36.031
TX_RESP_Q018		Merenda	0	30	3,1	2	3,6	36.032
TX_RESP_Q019		Manutenção	0	30	3,4	2	3,7	35.887
TX_RESP_Q020		Segurança	0	30	2,9	2	3,7	34.936
TX_RESP_Q021		Outras atividades	0	30	5,2	4	5,3	33.099

Ainda sobre as informações do diretor, as questões da Tabela 11 (TX_RESP_Q022 a TX_RESP_Q033 - Preparo para atividades) apresenta o nível de preparo para a realização de algumas atividades referentes ao cargo de diretor de uma escola. Vale ressaltar que para todas as atividades os diretores encontram-se em grande maioria preparados.

Tabela 11: TX_RESP_Q022 a TX_RESP_Q033 - Preparo para atividades

Questão	Quanto você sente estar preparado(a) para realizar a seguinte atividade:	Nada ou Pouco preparado	Preparado ou Muito preparado	Total Geral
TX_RESP_Q022	Liderar a equipe escolar.	3%	97%	37.363
TX_RESP_Q024	Atender as demandas administrativas da escola.	3%	97%	37.294
TX_RESP_Q031	Avaliar o desempenho dos(as) professores(as).	5%	95%	37.230
TX_RESP_Q023	Atender as demandas administrativas da rede escolar.	6%	94%	37.225
TX_RESP_Q029	Administrar conflitos.	6%	94%	37.216
TX_RESP_Q032	Realizar a autoavaliação institucional.	7%	93%	37.310
TX_RESP_Q033	Melhorar os processos pedagógicos da sua escola.	7%	93%	37.433
TX_RESP_Q030	Manter os(as) professores(as) motivados(as).	8%	92%	37.185
TX_RESP_Q025	Garantir a manutenção da escola.	9%	91%	37.242
TX_RESP_Q027	Mobilizar a comunidade para auxiliar a escola.	11%	91%	37.235
TX_RESP_Q028	Coordenar a implantação do Projeto Político-Pedagógico.	10%	90%	37.311
TX_RESP_Q026	Resolver as demandas dos familiares dos(as) alunos(as).	10%	90%	37.267

4.2.3 Condições de funcionamento da escola

Em cada questionário, há a seção sobre as condições gerais do funcionamento da escola. Inicialmente, com a Tabela 12, os diretores assinalaram as etapas educacionais presentes na sua respectiva escola. Os principais tipos de escolas são as que apresentam somente os anos iniciais do ensino fundamental e também a combinação dos anos iniciais somente com a pré-escola ou com os anos finais do ensino fundamental. Também é possível notar que das respostas válidas, 99% delas disseram possuir os anos iniciais do ensino fundamental na escola, restando 1% aparentemente sem turmas de 5ºano. Como as avaliações da SAEB são realizadas com os alunos dessa turma, existe algum nível de incoerência nessas respostas, as quais não serão eliminadas das outras análises, mas consideradas como uma margem de erro existente nos resultados.

Tabela 12: TX_RESP_Q034 a TX_RESP_Q038 - Etapas educacionais da escola

Educação Infantil e Creche (0 a 3 anos)	Educação Infantil e Pré-escola (4 e 5 anos)	Anos Iniciais do Ensino Fundamental	Anos Finais do Ensino Fundamental	Ensino Médio	Contagem	%
		X			8.930	24%
	X	X			7.795	21%
		X	X		7.993	21%
	X	X	X		5.028	13%
		X	X	X	2.568	7%
X	X	X	X		2.489	7%
X	X	X			2.266	6%
	X	X	X	X	317	1%
X	X	X	X	X	115	≈0%
			X		106	≈ 0%
	X				61	≈ 0%
		X		X	45	≈0%
X	X				38	≈0%
			X	X	33	≈0%
X		X	X		24	≈0%
X		X			23	≈0%
	X		X		14	≈0%
	X	X		X	8	≈0%
	X	X			7	≈0%
				X	6	≈0%
X	X	X		X	5	≈0%
X	X		X		3	≈0%
X		X	X	X	1	≈0%
X		X			1	≈0%
X			X		1	≈0%
X					1	≈0%
4.966	18.122	35.404	34.711	18.673	37.871	100%
13%	48%	99%	49%	8%	100%	

Para as questões *TX_RESP_Q039: A educação infantil funciona na sede da escola?* e *TX_RESP_Q040: A área externa é utilizada em horários diferenciados pelos(as) alunos(as) da Educação Infantil?* foram obtidas respostas positivas em 94% e 83% dos casos respectivamente.

Na Tabela 13 (TX_RESP_Q041 a TX_RESP_Q056 - Condições de funcionamento) nota-se o grau de concordância dos diretores sobre algumas afirmações, por exemplo, grande maioria acredita que a comunidade apoiou na gestão da escola, houve troca de experiências, pontualidade em início de aulas e apoio da Secretaria da Educação. No entanto, houve alta taxa de discordância em afirmações como "todos alunos receberam seus livros didáticos"e "os recursos financeiros disponíveis foram suficientes", o que pode ter afetado o decorrer do ano letivo.

Tabela 13: TX_RESP_Q041 a TX_RESP_Q056 - Condições de funcionamento

Questão	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano:	Discordo	%	Concordo	%	Total
TX_RESP_Q052	A comunidade apoiou a gestão da escola.	1.400	4%	36.301	96%	37.701
TX_RESP_Q051	Troquei experiências com diretores(as) de outras escolas.	1.994	5%	35.749	95%	37.743
TX_RESP_Q048	Os(As) professores(as) iniciaram as aulas no horário marcado.	2.693	7%	35.045	93%	37.738
TX_RESP_Q046	Recebi apoio da Secretaria de Educação.	4.228	11%	33.511	89%	37.739
TX_RESP_Q047	Os(As) professores(as) foram assíduos(as).	6.217	16%	31.485	84%	37.702
TX_RESP_Q050	Os(As) alunos(as) foram assíduos(as).	7.201	19%	30.531	81%	37.732
TX_RESP_Q042	Havia professores(as) para todas as disciplinas.	8.088	21%	29.659	79%	37.747
TX_RESP_Q054	As famílias contribuíram com o trabalho pedagógico.	12.023	32%	25.674	68%	37.697
TX_RESP_Q055	Os livros didáticos foram entregues antes do início das aulas.	12.295	33%	25.301	67%	37.596
TX_RESP_Q049	As substituições das ausências de professores(as) foram facilmente realizadas.	12.748	34%	25.001	66%	37.749
TX_RESP_Q044	Havia quantidade suficiente de pessoal para apoio pedagógico (coordenador, orientador etc.).	14.719	39%	23.046	61%	37.765
TX_RESP_Q043	Havia quantidade suficiente de pessoal administrativo.	15.403	41%	22.342	59%	37.745
TX_RESP_Q053	A comunidade executou trabalhos voluntários na escola.	15.845	42%	21.857	58%	37.702
TX_RESP_Q045	Os recursos pedagógicos foram suficientes.	17.387	46%	20.344	54%	37.731
TX_RESP_Q056	Todos(as) os(as) alunos(as) receberam livros didáticos.	19.849	53%	17.794	47%	37.643
TX_RESP_Q041	Os recursos financeiros foram suficientes.	24.852	66%	12.935	34%	37.787

Sobre o ano letivo, também há na Tabela 14 (TX_RESP_Q057 a TX_RESP_Q066 - Interrupções do calendário escolar) registros de casos de interrupções das aulas. Em 15% das instituições houve interrupções, dessas, há destaque para interrupções devido a greve de professores (41%) e eventos externos (40%). Para a questão *TX_RESP_Q067: Por quantos dias ocorreu a interrupção?* tem-se uma média de 11 dias e máximo de 60 dias.

Tabela 14: TX_RESP_Q057 a TX_RESP_Q066 - Interrupções do calendário escolar

Questão	Descrição		Não	%	Sim	%	Total
TX_RESP_Q057	O calendário escolar pré-estabelecido foi cumprido sem interrupções?		5.534	15%	32.156	85%	37.690
TX_RESP_Q061		Greve de professores	3.240	59%	2.294	41%	5.534
TX_RESP_Q064		Eventos externos	3.320	60%	2.214	40%	5.534
TX_RESP_Q066		Outros	3.945	71%	1.589	29%	5.534
TX_RESP_Q058		Falta de água	4.746	86%	788	14%	5.534
TX_RESP_Q063	Qual o motivo da interrupção do calendário?	Problemas de infraestrutura da escola	4.761	86%	773	14%	5.534
TX_RESP_Q065		Eventos climáticos (inundação, desmoronamento etc)	4.782	86%	752	14%	5.534
TX_RESP_Q059		Falta de energia	5.078	92%	456	8%	5.534
TX_RESP_Q060		Falta de merenda	5.358	97%	176	3%	5.534
TX_RESP_Q062		Vandalismo nas instalações	5.450	98%	84	2%	5.534

Por fim, na Tabela 15(TX_RESP_Q068 a TX_RESP_Q075 - Incidentes de segurança escolar) há o registro de que a maioria das escolas nunca sofreu nenhuma situação

do tipo. Entretanto, ainda que seja minoria, para casos de ameaças a profissionais por alunos, porte de arma, apresentação de aluno sob efeito de bebida alcoólica ou drogas ilícitas, há um número relevante - de 13% a 23% - de casos em que isso já ocorreu poucas vezes e de 1% a 2% de casos em que isso ocorre várias vezes. Para exemplificar a gravidades de tais número, em termos absolutos, a quantidade de escolas que normalmente já sofreu com ameaças a profissionais é de 606, por exemplo.

Tabela 15: TX_RESP_Q068 a TX_RESP_Q075 - Incidentes de segurança escolar

Questão	Sobre os fatos listados abaixo, diga a frequência com que ocorreram neste ano, nesta escola:	Nunca	%	Poucas vezes	%	Várias vezes	%	Total
TX_RESP_Q075	Episódios de violência ocasionaram cancelamento das aulas	36.479	97%	1.140	3%	92	0%	37.711
TX_RESP_Q070	Houve ocorrência de roubo com uso de violência	34.780	92%	2.646	7%	234	1%	37.660
TX_RESP_Q068	Profissionais foram vítimas de atentado à vida	34.411	91%	3.061	8%	203	1%	37.675
TX_RESP_Q071	Houve tráfico de drogas na escola	34.455	91%	2.915	8%	286	1%	37.656
TX_RESP_Q072	Alunos(as) frequentaram a escola sob efeito de bebida alcoólica	32.684	87%	4.755	13%	152	0%	37.591
TX_RESP_Q074	Alunos(as) frequentaram a escola portando arma (revólver, faca, canivete etc.)	32.323	86%	5.259	14%	80	0%	37.662
TX_RESP_Q073	Alunos(as) frequentaram a escola sob efeito de drogas ilícitas	31.662	84%	5.529	15%	426	1%	37.617
TX_RESP_Q069	Profissionais foram ameaçados(as) por algum aluno	28.262	75%	8.808	23%	606	2%	37.676

4.2.4 Recursos e infraestrutura

Nessa seção do questionário, aspectos das instalações físicas são descritas. No caso das escolas que atendem à educação básica (creche e/ou pré-escola), há espaço destinado à amamentação e armazenamento de leite em somente 3% e 2% das escolas, enquanto 55% delas possuem chuveiros para uso das crianças. Essas informações podem ser visualizadas na Tabela 16 (TX_RESP_Q076 a TX_RESP_Q078 - Instalações educação infantil).

Tabela 16: TX_RESP_Q076 a TX_RESP_Q078 - Instalações educação infantil

Questão	Descrição	Não	%	Sim	%	Total
TX_RESP_Q076	Há na escola espaço destinado exclusivamente à amamentação?	4.749	97%	138	3%	4.887
TX_RESP_Q077	Há na escola local para armazenamento de leite materno?	4.772	98%	118	2%	4.890
TX_RESP_Q078	Há na escola banheiro com chuveiro para uso das crianças?	7.853	45%	9.553	55%	17.406

Para a Tabela 17 (TX_RESP_Q079 a TX_RESP_Q102 - Aspectos físicos da escola), nota-se que as distribuições de elementos de área externa estão bem equilibrados em presença e não presença - há destaque para a alta presença de bebedouro à altura das crianças e a ausência de hortas. Sobre o revestimento do solo, também há uma boa distribuição, com exceção da ausência de escolas com piso emborrachado na área externa - o que poderia auxiliar na segurança das crianças ao brincar. Além disso, a maioria das escolas não possui os equipamentos de recreação elencados em sua área externa, principalmente

tanques de areia e túneis lúdicos.

Tabela 17: TX_RESP_Q079 a TX_RESP_Q102 - Aspectos físicos da escola

Questão	Descrição		Não	%	Sim	%	Total
TX_RESP_Q079		Bebedouro ao alcance das crianças	3.493	9%	34.232	91%	37.725
TX_RESP_Q082		Área coberta	10.008	27%	27.717	73%	37.725
TX_RESP_Q081	Sobre a área externa da sua escola	Área sombreada	14.462	38%	23.263	62%	37.725
TX_RESP_Q080	(pátio, parque e área verde), indique os itens existentes:	Banheiro infantil	16.317	43%	21.408	57%	37.725
TX_RESP_Q083		Vegetação e jardim	18.395	49%	19.330	51%	37.725
TX_RESP_Q084		Horta	26.361	70%	11.364	30%	37.725
TX_RESP_Q087		Cimento áspero	16.253	43%	21.472	57%	37.725
TX_RESP_Q090		Terra	19.534	52%	18.191	48%	37.725
TX_RESP_Q086		Cimento liso	22.537	60%	15.188	40%	37.725
TX_RESP_Q089	Quais os tipos de solo ou revestimento	Grama	23.054	61%	14.671	39%	37.725
TX_RESP_Q091	do solo da área externa da sua escola?	Cerâmica	27.227	72%	10.498	28%	37.725
TX_RESP_Q088		Areia	28.191	75%	9.534	25%	37.725
TX_RESP_Q092		Pedras	30.650	81%	7.075	19%	37.725
TX_RESP_Q085		Piso emborrachado	36.793	98%	932	2%	37.725
TX_RESP_Q101		Banco	24.300	64%	13.425	36%	37.725
TX_RESP_Q102		Outros	24.753	66%	12.972	34%	37.725
TX_RESP_Q097		Escorregador	26.964	71%	10.761	29%	37.725
TX_RESP_Q099		Balanço	28.422	75%	9.303	25%	37.725
TX_RESP_Q096	Quais equipamentos existem nas	Gangorra	30.467	81%	7.258	19%	37.725
TX_RESP_Q098	áreas externas de recreação da sua escola?	Casinha	31.965	85%	5.760	15%	37.725
TX_RESP_Q100		Brinquedo para escalar	32.960	87%	4.765	13%	37.725
TX_RESP_Q094		Gira-gira	33.341	88%	4.384	12%	37.725
TX_RESP_Q093		Tanque de areia	34.248	91%	3.477	9%	37.725
TX_RESP_Q095		Túnel lúdico	36.158	96%	1.567	4%	37.725

Para a Tabela 18 - TX_RESP_Q103 a TX_RESP_Q108 - Condições de acesso das áreas externas, há maior inadequação de acesso para os alunos da educação especial e também das condições de uso dos equipamentos de recreação. Sobre o acesso à entrada principal, a segurança na entrada e saída de alunos e a identificação do prédio como instituição escolar, acredita-se em maioria que encontram-se adequados. Essa situação pode ser problemática quanto à inclusão dos alunos e às condições de recreação as quais estão inseridos.

Tabela 18: TX_RESP_Q103 a TX_RESP_Q108 - Condições de acesso das áreas externas

Questão	Avalie os seguintes aspectos da escola:	Inadequado	%	Adequado	%	Total Geral
TX_RESP_Q104	O acesso à área externa de recreação pelos(as) alunos(as) público-alvo da educação especial.	20.055	57%	15.179	43%	35.234
TX_RESP_Q103	Condições de uso dos equipamentos da área externa de recreação.	17.620	53%	15.761	47%	33.381
TX_RESP_Q105	O acesso à entrada principal das pessoas com deficiência física e visual (ex.: rampas e marcadores no chão).	17.975	48%	19.330	52%	37.305
TX_RESP_Q106	Segurança na entrada e saída dos(as) alunos(as) da escola.	10.233	27%	27.356	73%	37.589
TX_RESP_Q107	Muros e grades que impedem que os(as) alunos(as) saiam sozinhos(as).	7.457	19%	30.177	81%	37.634
TX_RESP_Q108	Identificação externa que caracterize o prédio como uma instituição escolar.	6.875	18%	30.819	82%	37.694

Voltando à restrição das escolas que atendem ao ensino infantil (creche e/ou pré-escola), na Tabela 19 (TX_RESP_Q109 a TX_RESP_Q117 - Aquisição de bens de consumo), a aquisição de brinquedos, recursos pedagógicos e materiais de higiene pessoal

é feita de forma não exclusiva pela compra direta da escola ou Secretaria de Educação em pelo menos em 80% das escolas. Também há casos de recebimento de doações ou ainda solicitação para os responsáveis da criança matriculada.

Tabela 19: TX_RESP_Q109 a TX_RESP_Q117 - Aquisição de bens de consumo

Questão	Descrição		Não	%	Sim	%	Total
TX_RESP_Q109	Como a escola adquire Brinquedos:	Compra realizada pela própria escola/Secretaria de Educação	3.610	20%	14.579	80%	18.189
TX_RESP_Q110		Doações e campanhas de arrecadação	12.934	71%	5.255	29%	18.189
TX_RESP_Q111		Solicitado aos responsáveis pelas crianças	16.211	89%	1.978	11%	18.189
TX_RESP_Q112	Como a escola adquire Recursos pedagógicos:	Compra realizada pela própria escola/Secretaria de Educação	435	2%	17.754	98%	18.189
TX_RESP_Q113		Doações e campanhas de arrecadação	16.389	90%	1.800	10%	18.189
TX_RESP_Q114		Solicitado aos responsáveis pelas crianças	17.136	94%	1.053	6%	18.189
TX_RESP_Q115	Como a escola adquire Materiais de higiene pessoal:	Compra realizada pela própria escola/Secretaria de Educação	842	5%	17.347	95%	18.189
TX_RESP_Q116		Doações e campanhas de arrecadação	17.401	96%	788	4%	18.189
TX_RESP_Q117		Solicitado aos responsáveis pelas crianças	16.970	93%	1.219	7%	18.189

4.2.5 Gestão e participação

Para essa seção, foram discutidas questões acerca dos Conselhos escolares. Para a questão *TX_RESP_Q118: O Conselho Escolar é um colegiado geralmente constituído por representantes da escola e da comunidade que tem como objetivo acompanhar as atividades escolares. Na sua escola existe Conselho Escolar?* foram obtidas 83% de respostas positivas. Dessas, um aprofundamento sobre a periodicidade e composição do conselho estão na Tabela 20 - TX_RESP_Q119 a TX_RESP_Q124 - Composição do conselho escolar: há uma média de realização de 6 reuniões de conselho por ano, e uma composição média de 5 professores, 3 alunos, 4 pais e 3 funcionários.

Tabela 20: TX_RESP_Q119 a TX_RESP_Q124 - Composição do conselho escolar

Questão	Descrição		Mínimo	Máximo	Média	Mediana	Desvio padrão	Total
TX_RESP_Q119	Quantas reuniões do Conselho Escolar ocorreram neste ano?		1	20	6	4,0	3,6	31.007
TX_RESP_Q120	Considere os atores relacionados a seguir e indique quantos participam do Conselho Escolar:	Professores(as)	1	10	5	4,0	2,7	30.936
TX_RESP_Q121		Alunos(as)	1	10	3	2,0	2,2	19.778
TX_RESP_Q122		Pais (ou responsáveis)	1	10	4	3,0	2,3	30.712
TX_RESP_Q123		Funcionários	1	10	3	3,0	2,2	30.109
TX_RESP_Q124		Outros membros	1	10	3	2,0	2,0	19.141

Para os resultados da Tabela 21 (TX_RESP_Q125 a TX_RESP_Q128 - Temas de reuniões do conselho), pode-se dizer que os temas sobre questões financeiras e administrativas são tratados de forma mais recorrente que questões pedagógicas e de relacionamento com a comunidade.

Tabela 21: TX_RESP_Q125 a TX_RESP_Q128 - Temas de reuniões do conselho

Questão	Neste ano, indique a frequência com que os temas/assuntos foram discutidos pelo Conselho Escolar:	Nunca/Poucas vezes	%	Muitas vezes/Sempre	%	Total Geral
TX_RESP_Q125	Questões pedagógicas	11.443	36%	19.784	64%	31.227
TX_RESP_Q126	Questões administrativas e institucionais	8.812	28%	22.387	72%	31.199
TX_RESP_Q127	Questões financeiras	6.192	20%	24.949	80%	31.141
TX_RESP_Q128	Questões de relacionamento com a comunidade	12.810	41%	18.373	59%	31.183

Em relação à questão *TX_RESP_Q129 - O Conselho Escolar tem função deliberativa?* das 30.976 respostas, 97% foram positivas. Para as questões *TX_RESP_Q130 e TX_RESP_Q131 - O Conselho de Classe é um órgão formado por todos os professores que lecionam em cada turma/ano. Neste ano e nesta escola, quantas vezes se reuniu o Conselho de Classe?* das 37.871 respostas válidas, 17% afirmou não possuir conselho de classe. Dos 83% que possuem, a média de encontros no ano foi igual a 3.

Para as questões *TX_RESP_Q132 e TX_RESP_Q133 - A APM - Associação de Pais e Mestres existe para apoiar as ações da escola e integrar a comunidade. Neste ano e nesta escola, quantas vezes se reuniu a APM (ou caixa escolar)?* das 37.871 respostas válidas, 55% afirmou não possuir APM. Para os 45% restantes, há uma média de reuniões da APM igual a 4.

Para a questão *TX_RESP_Q134 - Há Grêmio Estudantil?* das 18.618 respostas válidas, somente 21% das escolas afirmou possuir um grêmio estudantil ativo, enquanto 55% não possui grêmio.

Na Tabela 22 (TX_RESP_Q135 a TX_RESP_Q137 - Tipo de administração escolar), é possível verificar que 99% das escolas respondentes não são administradas pela Polícia Militar, 87% não segue orientação religiosa e que há preparo para testes e avaliações externas em 92% das escolas.

Tabela 22: TX_RESP_Q135 a TX_RESP_Q137 - Tipo de administração escolar

Questão	Descrição	Não	%	Sim	%	Total
TX_RESP_Q135	A escola é administrada pela Polícia Militar	37.230	99%	485	1%	37.715
TX_RESP_Q136	Os(As) estudantes são preparados(as) para os testes de avaliação externos.	2.965	8%	34.735	92%	37.700
TX_RESP_Q137	A escola segue orientação religiosa	32.695	87%	4.883	13%	37.578

Já na Tabela 23 (TX_RESP_Q138 a TX_RESP_Q145 - Fontes de recursos), estão as informações sobre as fontes de recursos financeiros de cada instituição. Delas, 95% recebe do Programa Dinheiro Direto da Escola, 70% capta recursos em eventos promovidos

e 55% recebe repasses da rede de ensino, sendo essas as principais fontes de cada escola. Há menos destaque para contribuições voluntárias, ONGS e parcerias com empresas.

Tabela 23: TX_RESP_Q138 a TX_RESP_Q145 - Fontes de recursos

Questão	Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola:	Não	%	Sim	%	Total
TX_RESP_Q138	Programa Dinheiro Direto da Escola.	1.978	5%	35.702	95%	37.680
TX_RESP_Q139	Eventos promovidos nas dependências da escola (Festas, rifas etc.).	11.303	30%	26.402	70%	37.705
TX_RESP_Q142	Repasses da rede de ensino.	16.804	45%	20.459	55%	37.263
TX_RESP_Q145	Contribuições voluntárias dos(as) profissionais da escola.	25.184	67%	12.298	33%	37.482
TX_RESP_Q144	Contribuições voluntárias dos familiares dos(as) estudantes.	28.768	77%	8.672	23%	37.440
TX_RESP_Q141	Organizações sem fins lucrativos.	31.226	83%	6.210	17%	37.436
TX_RESP_Q140	Empresas que apoiam a escola.	32.704	87%	4.836	13%	37.540
TX_RESP_Q143	Pagamento de taxas pelos familiares dos(as) estudantes.	37.055	99%	445	1%	37.500

Em relação ao oferecimento de merenda, das 37.683 respostas válidas para a questão *TX_RESP_Q146 A escola oferece merenda aos(às) estudantes?* 99,5% das escolas afirmou que sim, há oferecimento. Para períodos de até 4h na escola, 63% das escolas oferece uma única refeição. A maioria das escolas não oferece períodos de permanência de 4h a 7h ou de mais de 7h, de modo que a quantidade de refeições para esses casos possui um universo de respostas reduzido, mas entende-se que há oferecimento de mais refeições aos alunos quanto maior o tempo na escola. Essas informações podem ser vistas na Tabela 24 - TX_RESP_Q147 a TX_RESP_Q149 - Oferecimento de merenda.

Tabela 24: TX_RESP_Q147 a TX_RESP_Q149 - Oferecimento de merenda

Questão	Quantas refeições são oferecidas nesta escola para alunos(as) que permanecem:	Não se aplica	%	Uma vez	%	Duas vezes	%	Três vezes ou mais	%	Total
TX_RESP_Q147	Menos de 4 horas na escola:	6.831	18%	23.199	63%	5.617	15%	1.408	4%	37.055
TX_RESP_Q148	Entre 4 e 7 horas na escola):	15.123	41%	6.866	19%	8.338	23%	6.305	17%	36.632
TX_RESP_Q149	Mais de 7 horas na escola:	28.211	78%	324	1%	1.663	5%	5.882	16%	36.080

Na Tabela 25 (TX_RESP_Q150 a TX_RESP_Q155 - Condições de oferecimento de merenda) estão os detalhes relacionados ao desempenho da escola no oferecimento das merendas. A maioria dos diretores respondentes concordou com todas es afirmações, havendo destaque positivo para a boa qualidade dos alimentos oferecidos e à quantidade de alimentos suficiente para os alunos da instituição. No entanto, apesar de ser a minoria das respostas, há representatividade em números absolutos para a falta de acesso de pessoas com mobilidade reduzida aos locais de alimentação, falta de acesso à pias de higienização e a alimentar-se sentado.

Tabela 25: TX_RESP_Q150 a TX_RESP_Q155 - Condições de oferecimento de merenda

Questão	Descrição	Concordo	%	Discordo	%	Total
TX_RESP_Q151	Os alimentos são de boa qualidade	35.844	96%	1.579	4%	37.423
TX_RESP_Q150	A quantidade de alimentos é suficiente para todos(as)	34.053	91%	3.407	9%	37.460
TX_RESP_Q152	A cozinha atende todas as necessidades do preparo da merenda	30.552	82%	6.843	18%	37.395
TX_RESP_Q154	O acesso ao local de alimentação é livre para pessoas com mobilidade reduzida	29.404	79%	7.938	21%	37.342
TX_RESP_Q153	Todos(as) conseguem se alimentar sentados	25.234	67%	12.193	33%	37.427
TX_RESP_Q155	Há pias para higienização das mãos próximas ao local de alimentação	24.057	64%	13.374	36%	37.431

Por fim, ainda sobre merendas, as questões *TX_RESP_Q156 - A merenda escolar é preparada na própria instituição?* e *TX_RESP_Q157 - Os cardápios da alimentação escolar são elaborados por nutricionista?* apresentam respostas afirmativas em 95% e 98% dos casos respectivamente, o que demonstra o cuidado com essa prestação de serviço dentro das escolas.

4.2.6 Gestão pedagógica

Nesta seção, há o esclarecimento de questões relacionadas ao projeto pedagógico da escolas, à formação de turmas, medidas de combate ao abandono escolar, entre outras.

Na Tabela 26 (TX_RESP_Q158 a TX_RESP_Q165 - Diretrizes do projeto político-pedagógico), é possível verificar que 94% das escolas possui um projeto político-pedagógico, desse montante, seu conteúdo amplamente discutido em 98% dos casos, e os professores e pais participam de sua elaboração na maioria das escolas. Nos projetos também há metas de aprendizagem, consideração de avaliações externas e alcance de indicadores para mais de 94% das instituições.

Tabela 26: TX_RESP_Q158 a TX_RESP_Q165 - Diretrizes do projeto político-pedagógico

Questão	Descrição	Não se aplica	%	Não	%	Sim	%	Total
TX_RESP_Q158	A escola possui Projeto Político-Pedagógico?	0	0%	2.100	6%	35.589	94%	37.689
TX_RESP_Q159	Seu conteúdo é discutido em reuniões?	157	0%	646	2%	34.740	98%	35.543
TX_RESP_Q160	Os(As) professores(as) participaram da elaboração?	95	0%	262	1%	35.179	99%	35.536
TX_RESP_Q161	Os pais participaram da elaboração?	971	3%	5.855	17%	28.632	81%	35.458
TX_RESP_Q162	Os(As) estudantes participaram da elaboração?	3.276	9%	10.126	29%	22.028	62%	35.430
TX_RESP_Q163	Estabelece metas de aprendizagem?	210	1%	542	2%	34.735	98%	35.487
TX_RESP_Q164	Considera os resultados de avaliações externas (SAEB, estaduais, municipais etc.)?	340	1%	757	2%	34.376	97%	35.473
TX_RESP_Q165	Há metas de alcance de indicadores externos (Ideb, índices estaduais ou municipais)?	453	1%	1.512	4%	33.507	94%	35.472

Na Tabela 27 (TX_RESP_Q166 a TX_RESP_Q198 - Critérios de matrícula e atribuição de turmas) é possível verificar que em 22% das escolas nem todas as solicitações de vaga resultaram em matrículas, para esses casos, precisaram ser utilizados critérios para

selecionar os estudantes em busca das vagas. Dos critérios, os principais utilizados foram a ordem de chegada/solicitação e o local de moradia da criança.

Já para os critérios de formação de turmas, os principais critérios empregados foram o agrupamento de alunos da mesma idade e a manutenção de turmas do ano anterior, enquanto o agrupamento por desempenho e afinidade entre estudantes foram pouco utilizados. Na alocação de turmas para os professores, a utilização de todos os critérios foi mais balanceada (desde preferências do professor até pela gestão da escola), com exceção da realização de sorteios, presente em somente 2% dos casos.

Ainda nessa tabela, verifica-se que existência de parcerias nas escolas dá-se principalmente com a Secretaria da Educação (96%) e Conselho Tutelar (87%), e, na minoria dos casos, com ONGs (21%).

Tabela 27: TX_RESP_Q166 a TX_RESP_Q198 - Critérios de matrícula e atribuição de turmas

Questão	Descrição		Não	%	Sim	%	Total
TX_RESP_Q166	Neste ano e nesta escola, todos que solicitaram vagas conseguiram se matricular?		8.434	22%	29.143	78%	37.577
TX_RESP_Q170	Neste ano, quais foram os critérios para a matrícula de novos estudantes nesta escola:	Ordem de chegada	3.779	45%	4.655	55%	8.434
TX_RESP_Q168		Local de moradia	4.567	54%	3.867	46%	8.434
TX_RESP_Q174		Outros	5.686	67%	2.748	33%	8.434
TX_RESP_Q167		Sorteio	8.136	96%	298	4%	8.434
TX_RESP_Q173		Desempenho do(a) estudante no ano anterior	8.089	96%	345	4%	8.434
TX_RESP_Q172		Características socioeconômicas	8.340	99%	94	1%	8.434
TX_RESP_Q169		Prova de seleção	8.399	100%	35	0%	8.434
TX_RESP_Q171		Cor ou raça do(a) estudante	8.432	100%	2	0%	8.434
TX_RESP_Q191	Quais critérios foram considerados para a formação das turmas:	Agrupar os(as) estudantes segundo a idade	21.508	57%	16.363	43%	37.871
TX_RESP_Q179		Manter as turmas existentes do ano anterior	23.396	62%	14.475	38%	37.871
TX_RESP_Q182		Outro	28.895	76%	8.976	24%	37.871
TX_RESP_Q178		Equilíbrio de meninos e meninas nas turmas	29.655	78%	8.216	22%	37.871
TX_RESP_Q175		Não se aplica	32.062	85%	5.809	15%	37.871
TX_RESP_Q177		Agrupar os(as) estudantes por critérios disciplinares	33.406	88%	4.465	12%	37.871
TX_RESP_Q181		Agrupar os(as) estudantes com base no seu desempenho	34.034	90%	3.837	10%	37.871
TX_RESP_Q176		Afinidade entre os(as) estudantes	36.003	95%	1.868	5%	37.871
TX_RESP_Q187	Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)?	Professores(as) experientes nas turmas com dificuldade de aprendizagem	16.091	43%	20.923	57%	37.014
TX_RESP_Q191		Atribuição pela gestão da escola	16.678	45%	20.269	55%	36.947
TX_RESP_Q185		Cursos de formação continuada realizados	17.070	46%	19.914	54%	36.984
TX_RESP_Q183		Preferência dos(as) professores(as)	18.815	50%	18.455	50%	37.270
TX_RESP_Q184		Tempo de serviço	19.237	52%	17.968	48%	37.205
TX_RESP_Q186		Professores(as) experientes nas turmas com facilidade de aprendizagem	21.735	59%	15.092	41%	36.827
TX_RESP_Q189		Revezamento dos(as) professores(as) entre séries/anos	21.614	59%	15.191	41%	36.805
TX_RESP_Q188		Manutenção do(a) professor(a) com a mesma turma	24.358	66%	12.538	34%	36.896
TX_RESP_Q190		Sorteio das turmas entre os(as) professores(as)	36.033	98%	624	2%	36.657
TX_RESP_Q195	Na sua escola há parcerias com:	Secretaria de Educação	1.537	4%	36.334	96%	37.871
TX_RESP_Q193		Conselho Tutelar	4.857	13%	33.014	87%	37.871
TX_RESP_Q194		Secretaria de Saúde	7.323	19%	30.548	81%	37.871
TX_RESP_Q196		Secretaria de Assistência Social	14.183	37%	23.688	63%	37.871
TX_RESP_Q197		Secretaria de Segurança Pública	21.233	56%	16.638	44%	37.871
TX_RESP_Q192		Ministério Público	22.906	60%	14.965	40%	37.871
TX_RESP_Q198		Organizações não governamentais/instituições privadas	29.833	79%	8.038	21%	37.871

Em relação à Tabela 28 (TX_RESP_Q199 a TX_RESP_Q205 - Mitigação de abandono e repetência escolares), é possível dizer que, para o abandono escolar, a medida mais eficiente, segundo os diretores, foi entrar em contato com os familiares do estudante, e, para a mitigação da repetência escolar, foi a revisão de práticas pedagógicas.

Tabela 28: TX_RESP_Q199 a TX_RESP_Q205 - Mitigação de abandono e repetência escolares

Questão	Descrição		Não foi realizada	%	Nada/Pouco efetiva	%	Efetiva/Muito efetiva	%	Total
TX_RESP_Q199	Neste ano, para redução do ABANDONO ESCOLAR, avalie o resultado das seguintes ações realizadas nesta escola:	Entrar em contato com os familiares do(a) estudante	1.132	3%	3.885	10%	32.750	87%	37.767
TX_RESP_Q200		Ir à residência do(a) estudante	7.567	20%	9.163	24%	20.983	56%	37.713
TX_RESP_Q201		Informar ao Conselho Tutelar	2.958	8%	6.851	18%	27.884	74%	37.693
TX_RESP_Q202	Neste ano, para a redução da REPETÊNCIA ESCOLAR, avalie o resultado das seguintes ações realizadas nesta escola:	Oferta de reforço escolar	3.947	10%	6.616	18%	27.122	72%	37.685
TX_RESP_Q203		Os(As) estudantes são estimulados(as) a apoiar uns(umas) aos(às) outros(as)	397	1%	4.161	11%	33.118	88%	37.676
TX_RESP_Q204		Revisão dos procedimentos de avaliação	370	1%	3.036	8%	34.248	91%	37.654
TX_RESP_Q205		Revisão das práticas pedagógicas	201	1%	2.730	7%	34.724	92%	37.655

Na Tabela 29 (TX_RESP_Q206 a TX_RESP_Q222 - Temáticas de projetos) é possível verificar os principais temas de projetos desenvolvidos nas escolas. Como destaque com alta presença, estão os projetos de educação ambiental - em 80% das escolas - e os projetos abordando o Bullying - em 78% das escolas. Com destaque para baixa presença, estão os projetos com temática de combate ao machismo - em apenas 13% das escolas - e de preservação dos direitos do idosos - em 17% das escolas. Essa distribuição pouco balanceada de temáticas poderia ser melhor compreendida - entender se existe, por exemplo, algum tipo de classificação de prioridade ou rotatividade de temas de acordo com o ano, período, composição docente, etc que justifique a alta variação.

Tabela 29: TX_RESP_Q206 a TX_RESP_Q222 - Temáticas de projetos

Questão	Nesta escola, há projetos com as seguintes temáticas	Não	%	Sim	%	Total
TX_RESP_Q213	Educação ambiental	7.444	20%	30.427	80%	37.871
TX_RESP_Q207	Bullying	8.430	22%	29.441	78%	37.871
TX_RESP_Q206	Violência	11.457	30%	26.414	70%	37.871
TX_RESP_Q217	Nutrição e alimentação	14.749	39%	23.122	61%	37.871
TX_RESP_Q211	Uso de drogas	14.890	39%	22.981	61%	37.871
TX_RESP_Q220	Direitos da criança e do adolescente	16.848	44%	21.023	56%	37.871
TX_RESP_Q218	Educação para o trânsito	16.992	45%	20.879	55%	37.871
TX_RESP_Q219	Relações étnico-raciais/racismo	19.242	51%	18.629	49%	37.871
TX_RESP_Q210	Sexualidade	22.575	60%	15.296	40%	37.871
TX_RESP_Q216	Desigualdades sociais	24.235	64%	13.636	36%	37.871
TX_RESP_Q214	Ciência e tecnologia	27.137	72%	10.734	28%	37.871
TX_RESP_Q222	Educação financeira e consumo sustentável	27.207	72%	10.664	28%	37.871
TX_RESP_Q221	Mundo do trabalho (direitos, relações etc)	28.818	76%	9.053	24%	37.871
TX_RESP_Q215	Diversidade religiosa	29.444	78%	8.427	22%	37.871
TX_RESP_Q209	Homofobia	30.322	80%	7.549	20%	37.871
TX_RESP_Q212	Direitos dos idosos	31.431	83%	6.440	17%	37.871
TX_RESP_Q208	Machismo	32.969	87%	4.902	13%	37.871

Por fim, nessa seção temos a Tabela 30 - TX_RESP_Q223 a TX_RESP_Q231 - Oferta de atividades de formação, na qual verifica-se que o formato em que mais escolas ofereceram atividades de formação para o corpo docente foi em avaliações de aprendizagem

- em 86% das escolas - e a menos presente foi a avaliação em larga escola - em 49% das escolas. Diferentes atividades de formação indicam que existe certo grau de diferença na metodologia de ensino empregada por cada escola e a preferência pelo método aplicado.

Tabela 30: TX_RESP_Q223 a TX_RESP_Q231 - Oferta de atividades de formação

Questão	Indique se neste ano a escola ofereceu atividades de formação nas seguintes áreas:	Não	%	Sim	%	Total
TX_RESP_Q224	Avaliação da aprendizagem.	9.605	14%	27.545	86%	37.150
TX_RESP_Q227	Conhecimento do currículo.	7.603	19%	29.608	81%	37.211
TX_RESP_Q226	Metodologias de ensino.	18.543	20%	18.067	80%	36.610
TX_RESP_Q229	Educação inclusiva.	12.070	25%	24.713	75%	36.783
TX_RESP_Q223	Conteúdo e compreensão dos conceitos da(s) área(s) de ensino.	27.207	26%	10.664	74%	37.871
TX_RESP_Q228	Gestão da sala de aula.	7.153	33%	30.022	67%	37.175
TX_RESP_Q231	Gestão e administração escolar.	18.491	50%	17.807	50%	36.298
TX_RESP_Q230	Novas tecnologias.	9.193	51%	27.896	49%	37.089
TX_RESP_Q225	Avaliação em larga escala.	5.287	51%	32.248	49%	37.535

4.2.7 Educação inclusiva

Nessa seção, a Tabela 31 (TX_RESP_Q232 a TX_RESP_Q251 - Componentes da educação inclusiva) apresenta informações sobre a capacidade das escolas de atender crianças especiais e de formas professores capazes de fazê-lo. Entende-se que em 48% das escolas a quantidade de profissionais de inclusão não são suficientes, sendo o posto de monitor de apoio o mais necessário para as escolas mal atendidas. Já em relação aos treinamentos oferecidos aos profissionais, 51% das escolas afirmou não ter recebido treinamentos para educação especial nos últimos doze meses. Das que receberam, a principal área de treinamento foi para Autismo e Transtornos Globais de Desenvolvimento (86%) e a que menos recebeu atenção foi a de Surdo-cegueira (14%). Esses números evidenciam o despreparo das escolas brasileiras para lidar do público alvo de atenção especial e a luta para transformação da inclusão em uma prática bem estruturada e eficiente.

Tabela 31: TX_RESP_Q232 a TX_RESP_Q251 - Componentes da educação inclusiva

Questão	Descrição		Não	%	Sim	%	Total
TX_RESP_Q232	Os(As) profissionais para inclusão do público-alvo da educação especial são em número suficiente?		18.051	48%	19.561	52%	37.612
TX_RESP_Q239	Indique quais outros(as) profissionais seriam necessários(as) no momento:	Monitor(a) de apoio	5.516	31%	12.535	69%	18.051
TX_RESP_Q234		Professor(a) de Educação Especial que atua na sala comum	7.080	39%	10.971	61%	18.051
TX_RESP_Q237		Professor(a) da sala de recursos multifuncionais	10.203	57%	7.848	43%	18.051
TX_RESP_Q238		Professor(a) itinerante	13.665	76%	4.386	24%	18.051
TX_RESP_Q240		Outros	14.343	79%	3.708	21%	18.051
TX_RESP_Q236		Professor(a) de Libras	14.598	81%	3.453	19%	18.051
TX_RESP_Q235		Professor(a) bilíngue para surdos	15.732	87%	2.319	13%	18.051
TX_RESP_Q233		Professor(a) de Braille	15.958	88%	2.093	12%	18.051
TX_RESP_Q241	Nos últimos doze meses, sua escola recebeu treinamento para lidar com o público-alvo da educação especial?		19.332	51%	18.359	49%	37.691
TX_RESP_Q244	Em quais áreas ocorreram treinamento:	Autismo ou outro Transtorno Global do Desenvolvimento	2.630	14%	15.729	86%	18.359
TX_RESP_Q242		Deficiência intelectual	3.569	19%	14.790	81%	18.359
TX_RESP_Q243		Deficiência física	10.486	57%	7.873	43%	18.359
TX_RESP_Q249		Deficiência múltipla	10.704	58%	7.655	42%	18.359
TX_RESP_Q248		Baixa visão	12.324	67%	6.035	33%	18.359
TX_RESP_Q246		Deficiência auditiva	12.344	67%	6.015	33%	18.359
TX_RESP_Q245		Surdez	12.849	70%	5.510	30%	18.359
TX_RESP_Q250		Altas habilidades/superdotação	13.345	73%	5.014	27%	18.359
TX_RESP_Q247		Cegueira	14.872	81%	3.487	19%	18.359
TX_RESP_Q251		Surdo-cegueira	15.803	86%	2.556	14%	18.359

4.3 Principais apontamentos da análise univariada

A partir da análise realizada com as informações da base consolidada é possível estabelecer um cenário resumo das condições das escolas brasileiras respondentes do questionário e participantes da SAEB 2019. Para cada subseção, as considerações sumarizadas serão apresentadas a seguir:

1. **Informações demográficas:** Destaque para a concentração de escolas nas regiões Nordeste e Sudeste e desiquilíbrio em quantidade de escolas por km^2 em diferentes estados do país. Há mais escolas em áreas urbanas e mais escolas em níveis socioeconômicos intermediários - níveis extremos baixos concentram-se no Norte e Nordeste, enquanto extremos altos no Sul e Sudeste. A taxa de participação dos alunos na avaliação da SAEB foi acima de 90% para todos estados. Para português, a média das avaliações foi de 206,06 - nível considerado proficiente, para matemática, a média foi de 219,65 - nível considerado básico.

2. **Informações do diretor:** Destaque para as altas médias de experiência prévia dos diretores respondentes como professores ou ainda como diretores em outras instituições, há também boa distribuição das horas de trabalho semanais nas diferentes funções e tarefas, sem concentração clara. Também vale ressaltar que a maioria dos diretores sente-se preparado para a realização de tais atividades.

3. **Condições de funcionamento da escola:** As escolas brasileiras possuem condições de funcionamento principalmente favoráveis enquanto apoio da comunidade, troca de experiências e pontualidade de atividades, no entanto, afirmam que os recursos financeiros, pedagógicos e o recebimento de livros não foi suficiente em sua maioria. Outro destaque é que 15% das escolas apresentou interrupções no calendário escolar, principalmente por greves e eventos externos de, em média, 11 dias. Outro ponto importante é a existência de incidentes de segurança escolar como ameaças ao corpo docente, porte de arma nas localidades da escola, entre outros que ocorreram em até 23% das instituições.

4. **Recursos e infraestrutura:** Há uma distribuição equilibrada entre presença e não-presença dos elementos de área externa e revestimento do solo. Nas condições de acesso às áreas externas, há maior inadequação de acesso para os alunos da educação especial e também das condições de uso dos equipamentos de recreação.

5. **Gestão e participação:** Existe Conselho Escolar em 83% das escolas respondentes, sua composição principal é de professores e pais, e os temas sobre questões financeiras e administrativas são tratados de forma mais recorrente. Também há maioria de escolas sem grêmio estudantil, escolas não administradas pela Polícia Militar e sem orientação religiosa. As principais fontes de recursos são o Programa Dinheiro Direto da Escola e em eventos promovidos. Por fim, sobre a oferta de merenda, a situação é favorável, em que 99,5% das escolas a oferecem aos alunos e afirmam possuir alimentos de qualidade e em quantidade suficiente.

6. **Gestão pedagógica:** É possível verificar que 94% das escolas possui um projeto político-pedagógico, desse montante, seu conteúdo amplamente discutido em 98% dos casos, e os professores e pais participam de sua elaboração na maioria das escola. Também destaca-se que a existência de parcerias nas escolas dá-se principalmente com a Secretaria da Educação (96%) e Conselho Tutelar (87%). Por fim, como temáticas de projeto destaque com alta presença, estão os projetos de educação ambiental - em 80% das escolas e os projetos abordando o Bullying - em 78% das escolas.

7. **Educação inclusiva:** Compreende-se que, em 48% das escolas, o número de profissionais de inclusão não é suficiente, com o cargo de monitor de apoio sendo o mais

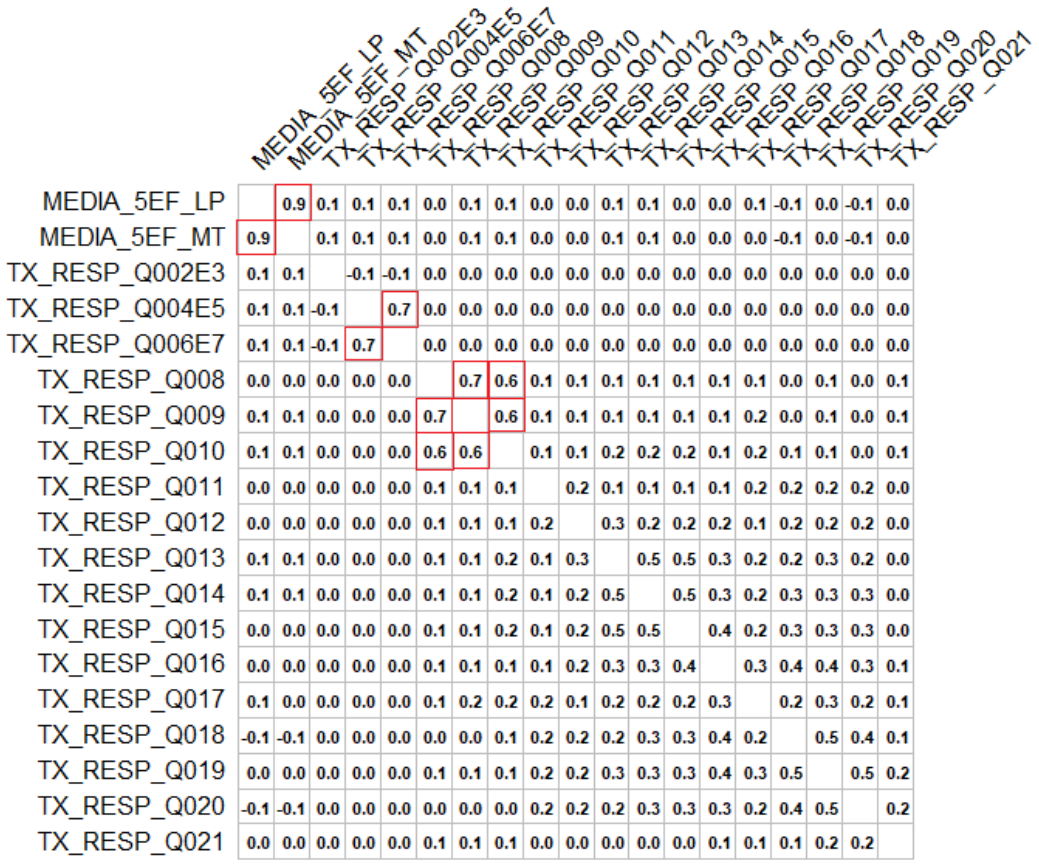
necessário para as escolas que estão carentes de pessoal. Além disso, em relação aos treinamentos oferecidos aos profissionais, 51% das escolas informaram não ter recebido capacitação para educação especial nos últimos doze meses.

4.4 Identificação de relação entre variáveis

Após a análise descritiva da base, foram construídas matrizes de correlação entre as variáveis contínuas do estudo utilizando o coeficiente de correlação de Pearson, com o objetivo de identificar e remover possíveis variáveis altamente correlacionadas, que poderiam aumentar o risco de sobreajuste do modelo. Isso pode ser visto na Figura 6 - Matriz de correlação - questões 02 a 21 e na Figura 7 - Matriz de correlação - questões 119 a 132. No entanto, os resultados das correlações não foram fortes o suficiente para justificar a exclusão de qualquer variável com esse método. Há exceção da substituição das variáveis MEDIA_5EF_LP e MEDIA_5EF_MT por uma nova variável de saída chamada MEDIA_5EF, justificada pela alta correlação entre as duas notas e que passou a ser representada pela média simples das notas de Língua Portuguesa (LP) e Matemática (MAT) - combinando os dois resultados em uma única métrica simplificada de desempenho dos alunos e escolas.

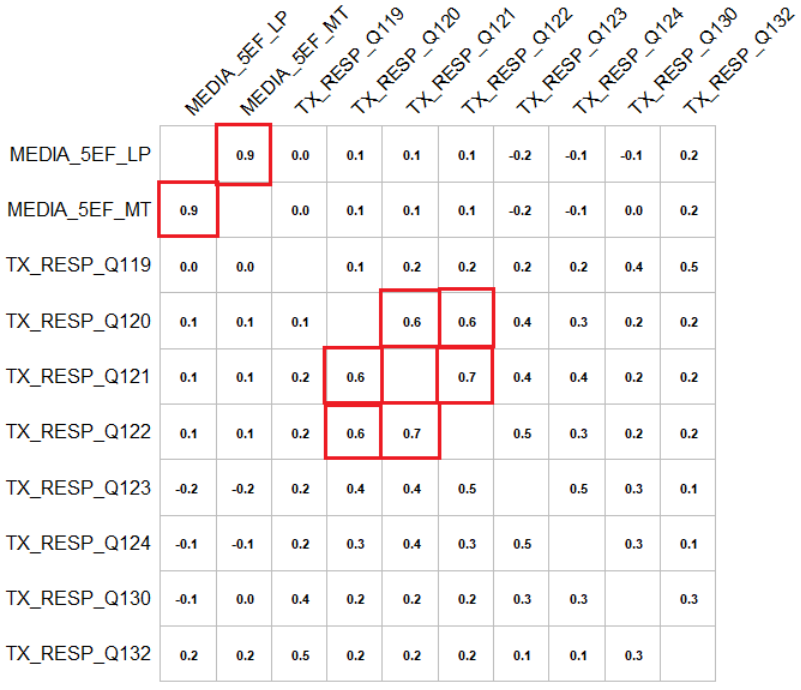
Após isso, foram retiradas da base também as questões com regra de dependência que afetassem o tamanho da amostra de registro, resultando em 203 variáveis aptas à entrada no modelo das árvores de regressão, 1 variável saída (MEDIA_5EF) e uma coluna de identificação (ID_ESCOLA). O relacional de todas as variáveis que irão compor os modelos e sua respectiva descrição pode ser encontrado no Anexo A - Lista Geral de Variáveis.

Figura 6: Matriz de correlação - questões 02 a 21



	MEDIA_5EF_LP	MEDIA_5EF_MT	TX_RESP_Q002E3	TX_RESP_Q004E5	TX_RESP_Q006E7	TX_RESP_Q008	TX_RESP_Q009	TX_RESP_Q010	TX_RESP_Q011	TX_RESP_Q012	TX_RESP_Q013	TX_RESP_Q014	TX_RESP_Q015	TX_RESP_Q016	TX_RESP_Q017	TX_RESP_Q018	TX_RESP_Q019	TX_RESP_Q020	TX_RESP_Q021
MEDIA_5EF_LP		0.9	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.1	-0.1	0.0	-0.1	0.0
MEDIA_5EF_MT	0.9		0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0
TX_RESP_Q002E3	0.1	0.1		-0.1	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TX_RESP_Q004E5	0.1	0.1	-0.1		0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TX_RESP_Q006E7	0.1	0.1	-0.1	0.7		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TX_RESP_Q008	0.0	0.0	0.0	0.0	0.0		0.7	0.6	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.0	0.1
TX_RESP_Q009	0.1	0.1	0.0	0.0	0.0	0.7		0.6	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.0	0.1	0.0	0.1
TX_RESP_Q010	0.1	0.1	0.0	0.0	0.0	0.6	0.6		0.1	0.1	0.2	0.2	0.2	0.1	0.2	0.1	0.1	0.0	0.1
TX_RESP_Q011	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1		0.2	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.0
TX_RESP_Q012	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.2		0.3	0.2	0.2	0.2	0.1	0.2	0.2	0.2	0.0
TX_RESP_Q013	0.1	0.1	0.0	0.0	0.0	0.1	0.1	0.2	0.1	0.3		0.5	0.5	0.3	0.2	0.2	0.3	0.2	0.0
TX_RESP_Q014	0.1	0.1	0.0	0.0	0.0	0.1	0.1	0.2	0.1	0.2	0.5		0.5	0.3	0.2	0.3	0.3	0.3	0.0
TX_RESP_Q015	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.1	0.2	0.5	0.5		0.4	0.2	0.3	0.3	0.3	0.0
TX_RESP_Q016	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.2	0.3	0.3	0.4		0.3	0.4	0.4	0.3	0.1
TX_RESP_Q017	0.1	0.0	0.0	0.0	0.0	0.1	0.2	0.2	0.2	0.1	0.2	0.2	0.2	0.3		0.2	0.3	0.2	0.1
TX_RESP_Q018	-0.1	-0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.2	0.2	0.3	0.3	0.4	0.2		0.5	0.4	0.1
TX_RESP_Q019	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.2	0.2	0.3	0.3	0.3	0.4	0.3	0.5		0.5	0.2
TX_RESP_Q020	-0.1	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.3	0.3	0.3	0.2	0.4	0.5		0.2
TX_RESP_Q021	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.2	0.2	

Figura 7: Matriz de correlação - questões 119 a 132



	MEDIA_5EF_LP	MEDIA_5EF_MT	TX_RESP_Q119	TX_RESP_Q120	TX_RESP_Q121	TX_RESP_Q122	TX_RESP_Q123	TX_RESP_Q124	TX_RESP_Q130	TX_RESP_Q132
MEDIA_5EF_LP		0.9	0.0	0.1	0.1	0.1	-0.2	-0.1	-0.1	0.2
MEDIA_5EF_MT	0.9		0.0	0.1	0.1	0.1	-0.2	-0.1	0.0	0.2
TX_RESP_Q119	0.0	0.0		0.1	0.2	0.2	0.2	0.2	0.4	0.5
TX_RESP_Q120	0.1	0.1	0.1		0.6	0.6	0.4	0.3	0.2	0.2
TX_RESP_Q121	0.1	0.1	0.2	0.6		0.7	0.4	0.4	0.2	0.2
TX_RESP_Q122	0.1	0.1	0.2	0.6	0.7		0.5	0.3	0.2	0.2
TX_RESP_Q123	-0.2	-0.2	0.2	0.4	0.4	0.5		0.5	0.3	0.1
TX_RESP_Q124	-0.1	-0.1	0.2	0.3	0.4	0.3	0.5		0.3	0.1
TX_RESP_Q130	-0.1	0.0	0.4	0.2	0.2	0.2	0.3	0.3		0.3
TX_RESP_Q132	0.2	0.2	0.5	0.2	0.2	0.2	0.1	0.1	0.3	

5 RESULTADOS DA MODELAGEM

Este capítulo apresenta os resultados obtidos nas etapas de modelagem e avaliação do estudo, partindo da estruturação do modelo, rodadas de teste até a consideração das árvores de regressão finais e respectivas interpretações. Cabe ressaltar que as tabelas e relações a seguir tem como fonte a elaboração própria. O código implementado para a criação das árvores pode ser encontrado na íntegra no Anexo B - Código Geral em R.

5.1 Composição e critérios para poda das árvores de regressão

Para a criação das árvores de regressão optou-se pelo agrupamento das variáveis da base em conjuntos temáticos que pudessem apresentar saídas melhor interpretáveis, além da geração de uma árvore única contendo todas as variáveis como cenário base. Entende-se que uma das premissas desse modelo seja entender a relação das variáveis com a nota, e não só descrever sua disposição com as informações atuais. Dessa forma, serão apresentadas as seguintes árvores para a variável única dependente MEDIA_5EF:

- **Árvore Geral**: composta por todas as 203 variáveis independentes da base apresentadas no Anexo A;
- **Árvore Geral Questionário Diretor**: composta pelas 198 variáveis advindas do "Questionário do Diretor", excluem-se as variáveis demográficas de localização das escolas das 203 variáveis totais;
- **Árvore Temática 1**: composta pelas 30 variáveis da seção "Informações do diretor";
- **Árvore Temática 2**: composta pelas 30 variáveis da seção "Condições de funcionamento da escola"e pelas 30 variáveis da seção "Recursos e infraestrutura";
- **Árvore Temática 3**: composta pelas 40 variáveis da seção "Gestão e participação", pelas 66 variáveis da seção "Gestão pedagógica:"e pelas 2 variáveis da seção "Educação inclusiva".

Vale ressaltar que diferentes combinações e árvores foram testadas e essa estrutura foi a que pôde direcionar melhor o estudo. Além disso, uma árvore própria para as 5 variáveis da seção "Informações demográficas"não foi criada, pois essas informações

não são características da escola, mas principalmente as posicionam no território e não objetiva-se compreender exclusivamente seu efeito sem o efeito das demais. Também porque o reflexo dessas variáveis está já explícito na Árvore Geral - na qual as informações demográficas são as mais importantes da base - e nas combinações de todas as árvores temáticas em conjunto com as informações demográficas, em que as informações demográficas sobrepuseram-se às demais em todos casos. Ou seja, se as outras árvores fossem construídas com as informações de localidade, os resultados seriam todos iguais, com a sobreposição da localização sempre e isso prejudicaria o objetivo do trabalho.

Já para a poda de cada árvore utilizaram-se como parâmetros do pacote *Rpart* e seus respectivos valores finais (vale ressaltar que a poda acontece assim que algum desses critérios de parada seja alcançado):

- **cp (Complexity Parameter) = 0.001**

  Esse valor define o nível mínimo de melhoria na divisão que justifica o crescimento da árvore. Valores baixos resultam em árvores maiores, enquanto valores mais altos favorecem a poda, reduzindo a complexidade. Esse valor foi escolhido com base no gráfico de complexidade da árvore (plotagem do cp), onde o erro de validação já se apresentava estabilizado (*flat*) ao redor de 0.001. Embora não seja o ponto ótimo, esse valor reduz a complexidade da árvore sem prejudicar o desempenho, pois o erro de validação não melhora significativamente para valores menores de cp.

- **minsplit = 1000**

  Esse parâmetro define o número mínimo de observações necessárias para que um nó seja dividido. Ao definir minsplit = 1000, evita-se que a árvore faça divisões em subconjuntos pequenos, o que poderia levar ao superajuste. Esse valor foi selecionado para garantir que as divisões ocorram apenas em grupos de dados suficientemente grandes, contribuindo para a estabilidade do modelo.

- **minbucket = 1000**

  O minbucket define o número mínimo de observações que um nó terminal pode conter - o valor de 1000 foi utilizado para promover um ajuste mais generalizável e evitar que a árvore crie nós finais com um número muito pequeno de amostras, o que poderia tornar o modelo sensível a variações específicas dos dados de treino.

- **maxdepth = 6**

  O parâmetro maxdepth limita a profundidade máxima da árvore, ajustar maxdepth a 6 níveis foi uma forma de controlar a complexidade. Árvores muito profundas tendem a se ajustar demais aos dados de treino, enquanto essa profundidade se mostrou suficiente para capturar padrões significativos sem sobreajustar.

- **xval = 10**

  A validação cruzada com 10 divisões (xval = 10) foi escolhida para avaliar o desempenho do modelo em diferentes subconjuntos de dados e assegurar que ele está generalizando adequadamente, dessa forma, é possível identificar se o modelo está ajustado demais aos dados de treino ou se está performando bem em dados ainda não vistos.

5.2 Árvores e interpretação

As árvores foram criadas com base em 80% dos dados - conjunto de treino do modelo - como havia sido mencionado na Seção 3 - Métodos e Premissas. O restante dos dados será utilizado para calcular o erro e realizar a validação do modelo, sendo uma etapa posterior à essa. Para cada árvore gerada, serão apresentadas a plotagem da árvore e o esquema da árvore em formato de texto, além da sua interpretação em termos práticos, retomando as questões das variáveis envolvidas na árvore.

5.2.1 Árvore Geral

Para a árvore geral, em que todas as 203 variáveis são incluídas no modelo, nota-se a prevalência das variáveis de caracterização demográficas das escolas, sendo a localização estadual (ID_UF), o nível socioeconômico, a região da escola e o tipo de localização (rural/urbano) os pontos mais relevantes para distribuição das notas. Além da visualização esquemática da árvore - que pode ser vista na Figura 8 - essas variáveis podem ser vistas na Figura - 9, em que está o esquema da árvore em formato de texto. O número da hierarquia de cada nó, de acordo com a ordem das divisões do algoritmo, está nessas imagens.

Para compreender o comportamento do desempenho, inicia-se a visualização pelo nó raiz ID_UF, o qual subdivide-se em outros dois nós de nível socio-econômico e assim

por diante até chegarmos aos nós terminais, suas previsões e a quantidade n de amostras que pertence a essa região de previsão. Para a árvore criada, os maiores desempenhos encontram-se no último nó à direita, (nota 241, com 6% das escolas no nó) e o menor para o primeiro nó a esquerda (nota 175, com 5% das escolas no nó).

Resumindo as condições, nota-se que escolas situadas em estados majoritariamente das regiões Norte, Nordeste e Centro-Oeste e que possuem nível socioeconômico mais baixos possuem um desempenho inferior quando comparadas ao restante da base.

Figura 8: Árvore de Regressão: Árvore Geral

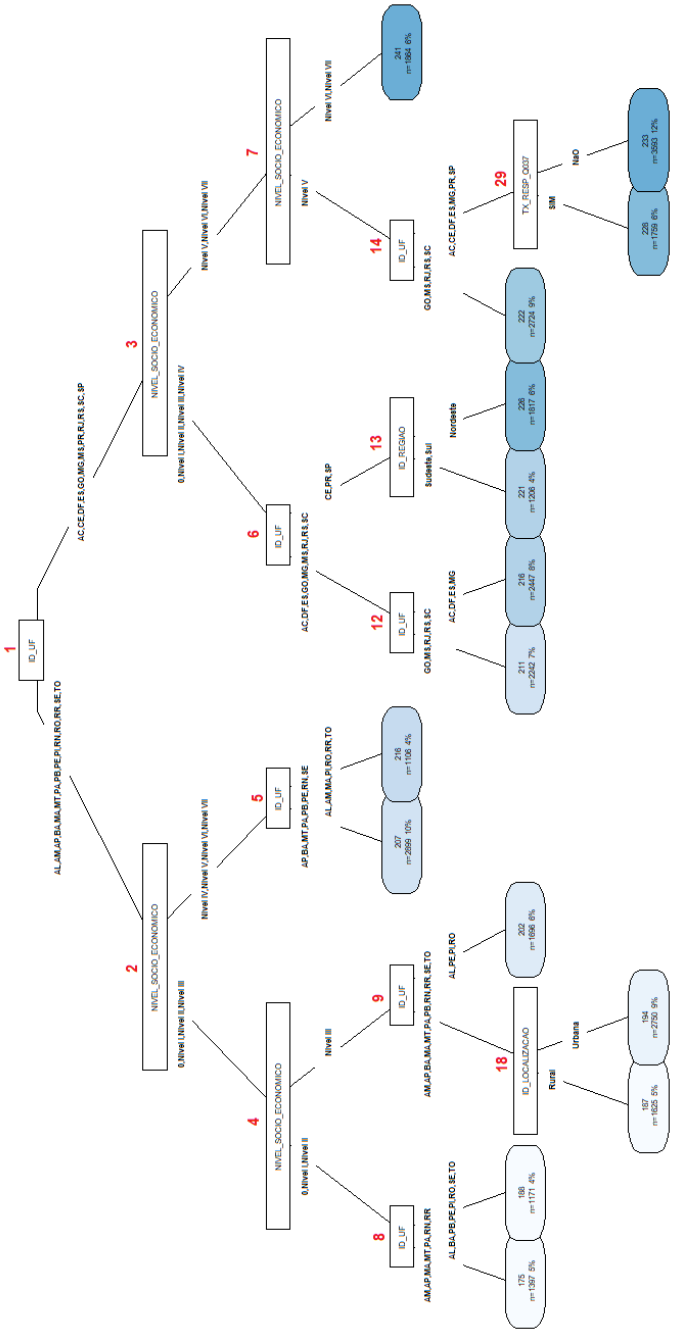


Figura 9: Print da Árvore Geral em formato de texto

```
n= 30296

node), split, n, deviance, yval
      * denotes terminal node

 1) root 30296 18656580.0 213.0412
   2) ID_UF=AL,AM,AP,BA,MA,MT,PA,PB,PE,PI,RN,RO,RR,SE,TO 12644  6431751.0 196.4838
     4) NIVEL_SOCIO_ECONOMICO=0,Nivel I,Nivel II,Nivel III 8639  3926753.0 190.3154
       8) NIVEL_SOCIO_ECONOMICO=0,Nivel I,Nivel II 2568  1277225.0 180.8772
        16) ID_UF=AM,AP,MA,MT,PA,RN,RR 1397   478591.1 174.6111 *
        17) ID_UF=AL,BA,PB,PE,PI,RO,SE,TO 1171   678342.9 188.3527 *
       9) NIVEL_SOCIO_ECONOMICO=Nivel III 6071  2324007.0 194.3078
        18) ID_UF=AM,AP,BA,MA,MT,PA,PB,RN,RR,SE,TO 4375  1355865.0 191.2886
          36) ID_LOCALIZACAO=Rural 1625   567719.2 187.1196 *
          37) ID_LOCALIZACAO=Urbana 2750   743213.3 193.7521 *
        19) ID_UF=AL,PE,PI,RO 1696   825384.8 202.0961 *
     5) NIVEL_SOCIO_ECONOMICO=Nivel IV,Nivel V,Nivel VI,Nivel VII 4005  1467270.0 209.7893
      10) ID_UF=AP,BA,MT,PA,PB,PE,RN,SE 2899   878564.7 207.3710 *
      11) ID_UF=AL,AM,MA,PI,RO,RR,TO 1106   527315.0 216.1279 *
   3) ID_UF=AC,CE,DF,ES,GO,MG,MS,PR,RJ,RS,SC,SP 17652  6275591.0 224.9012
     6) NIVEL_SOCIO_ECONOMICO=0,Nivel I,Nivel II,Nivel III,Nivel IV 7712  3096418.0 217.7456

      12) ID_UF=AC,DF,ES,GO,MG,MS,RJ,RS,SC 4689  1352652.0 213.6143
        24) ID_UF=GO,MS,RJ,RS,SC 2242   549919.5 210.6286 *
        25) ID_UF=AC,DF,ES,MG 2447   764436.8 216.3498 *
      13) ID_UF=CE,PR,SP 3023  1539597.0 224.1537
        26) ID_REGIAO=Sudeste,Sul 1206   267681.3 220.8050 *
        27) ID_REGIAO=Nordeste 1817  1249414.0 226.3764 *
     7) NIVEL_SOCIO_ECONOMICO=Nivel V,Nivel VI,Nivel VII 9940  2477935.0 230.4529
      14) NIVEL_SOCIO_ECONOMICO=Nivel V 8076  1829537.0 228.0440
        28) ID_UF=GO,MS,RJ,RS,SC 2724   631000.4 222.4668 *
        29) ID_UF=AC,CE,DF,ES,MG,PR,SP 5352  1070678.0 230.8827
          58) TX_RESP_Q037=SIM 1759   348662.4 227.5628 *
          59) TX_RESP_Q037=NaO 3593   693137.8 232.5080 *
      15) NIVEL_SOCIO_ECONOMICO=Nivel VI,Nivel VII 1864   398505.3 240.8895 *
```

5.2.2 Árvore Geral Questionário Diretor

Além da visualização esquemática da árvore - que pode ser vista na Figura 10 - as variáveis envolvidas podem ser vistas na Figura - 11, em que está o esquema da árvore em formato de texto. Para a árvore geral do questionário diretor, em que as 198 variáveis são incluídas no modelo, com exceção das variáveis de localização, nota-se a prevalência das seguintes variáveis e seus respectivos efeitos dentro de seus nós (a ordem de leitura dos nós de todas as árvores a seguir dá-se da esquerda a direita e em profundidade, exaurindo-se um ramo até seu nó terminal para voltar a níveis mais altos):

- TX_RESP_Q001 - "Qual é a sua cor ou raça?"em que a localização de escolas com maior desempenho está à direita, nível da cor branca.
- TX_RESP_Q158 - "A escola possui Projeto Político-Pedagógico?"para a ausência de projeto, o efeito é negativo no desempenho.
- TX_RESP_Q097- "Quais equipamentos existem nas áreas externas de recreação da sua escola? Escorregador"para a ausência do brinquedo, há um menor desempenho.

- TX_RESP_Q166 - "Neste ano e nesta escola, todos que solicitaram vagas conseguiram se matricular?", de forma pouco intuitiva, caso todos tenham conseguido se matricular, há um efeito negativo.

- TX_RESP_Q184 - "Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Tempo de serviço."para o caso do uso desse critério, há um efeito positivo no desempenho.

- TX_RESP_Q134 - "Há Grêmio Estudantil?"para a existência de grêmio, há um aumento das notas e desempenho das escolas.

- TX_RESP_Q144 - "Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Contribuições voluntárias dos familiares dos(as) estudantes."caso haja fornecimento dessa fonte, escolas apresentam maior desempenho.

- TX_RESP_Q083 - "Sobre a área externa da sua escola (pátio, parque e área verde), indique os itens existentes: Vegetação e jardim "para a existência de jardim, há efeito positivo.

- TX_RESP_Q037 - "Indique quais são as etapas educacionais atendidas pela sua escola: Anos Finais do Ensino Fundamental."para escolas que não atendem a essa etapa de ensino, há um melhor desempenho.

- TX_RESP_Q054 - "Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: As famílias contribuíram com o trabalho pedagógico."caso haja concordância com a contribuição das famílias, o desempenho é maior.

- TX_RESP_Q089 - "Quais os tipos de solo ou revestimento do solo da área externa da sua escola? Grama"para casos de uso de grama, há efeito positivo.

- TX_RESP_Q145 - "Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Contribuições voluntárias dos(as) profissionais da escola."para esse tipo de contribuição, caso exista, há efeito negativo no desempenho.

De forma geral, a existência de um projeto político pedagógico, grêmio, brinquedos de recreação e jardim/grama levam a melhores desempenhos, caracterizando uma escolas melhor preparada e equipada. Também nota-se que contribuições de responsáveis pelos alunos são potenciais positivos, enquanto oferecimento de recursos pelos próprios funcionários da instituição é algo negativo. Por fim, uma questão importante é a raça do diretor da escola. Pode-se compreender que o efeito no desempenho advindo de diretores de cor branca seja na verdade um reflexo de desigualdades da sociedade e não uma relação de causa e consequência, valendo a investigação dos diferentes contextos sociais em que essas escolas encontram-se.

Figura 10: Árvore de Regressão: Árvore Geral Questionário Diretor

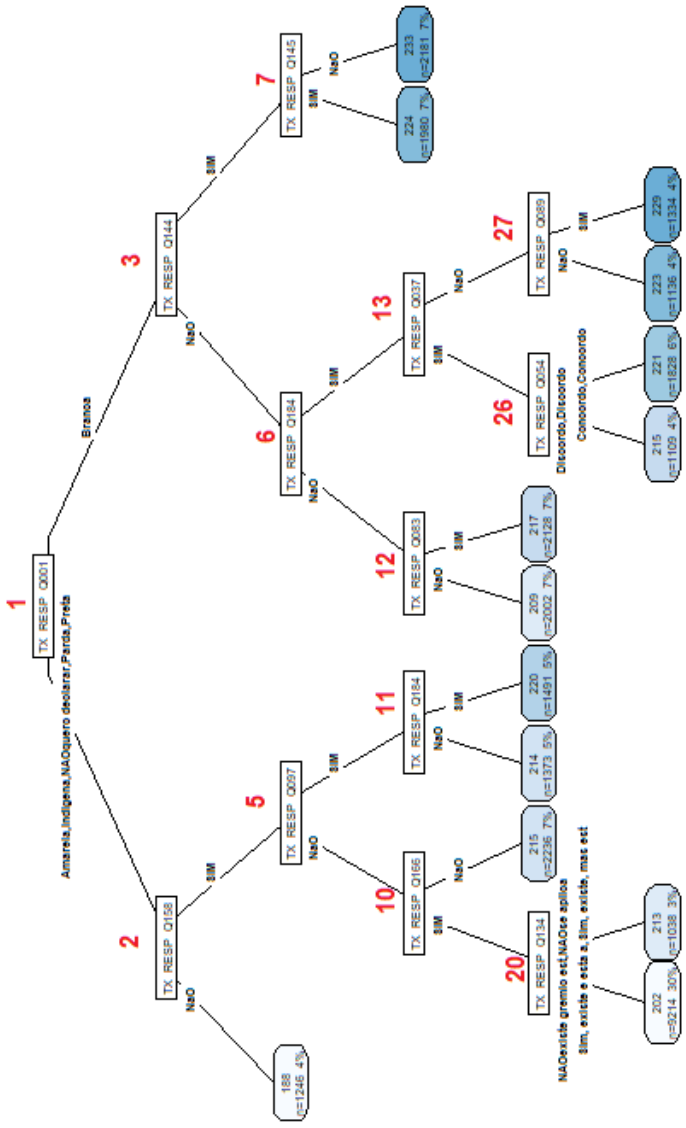


Figura 11: Print da Árvore Geral Questionário Diretor em formato de texto

```
n= 30296

node), split, n, deviance, yval
      * denotes terminal node

 1) root 30296 18767980.0 212.9454
   2) TX_RESP_Q001=Amarela,Indigena,NAOquero declarar,Parda,Preta 16598 10682420.0 206.0247
     4) TX_RESP_Q158=NaO 1246   839215.4 188.2482 *
     5) TX_RESP_Q158=SIM 15352  9417506.0 207.4674
      10) TX_RESP_Q097=NaO 12488  7742449.0 205.2216
        20) TX_RESP_Q166=SIM 10252  6543967.0 203.1723
          40) TX_RESP_Q134=NAOexiste gremio estudantil,NAOse aplica 9214  5700129.0 202.1204 *
          41) TX_RESP_Q134=Sim, existe e esta ativo,Sim, existe, mas esta inativo 1038   743141.9 212.5097 *
        21) TX_RESP_Q166=NaO 2236   958020.7 214.6177 *
      11) TX_RESP_Q097=SIM 2864  1337436.0 217.2599
        22) TX_RESP_Q184=NaO 1373   765839.2 213.8548 *
        23) TX_RESP_Q184=SIM 1491   541017.9 220.3955 *
   3) TX_RESP_Q001=Branca 13698  6327288.0 221.3313
     6) TX_RESP_Q144=NaO 9537  4494933.0 218.2763
      12) TX_RESP_Q184=NaO 4130  2224590.0 213.1150
        24) TX_RESP_Q083=NaO 2002  1081070.0 208.7054 *
        25) TX_RESP_Q083=SIM 2128  1067969.0 217.2634 *
      13) TX_RESP_Q184=SIM 5407  2076288.0 222.2186
        26) TX_RESP_Q037=SIM 2937  1103665.0 218.7013
          52) TX_RESP_Q054=Discordo,Discordo  1109   430146.6 215.2780 *
          53) TX_RESP_Q054=Concordo,Concordo  1828   652636.6 220.7782 *
        27) TX_RESP_Q037=NaO 2470   893084.9 226.4009
          54) TX_RESP_Q089=NaO 1136   478735.2 223.3130 *
          55) TX_RESP_Q089=SIM 1334   394294.0 229.0304 *
     7) TX_RESP_Q144=SIM 4161  1539331.0 228.3334
      14) TX_RESP_Q145=SIM 1980   903776.2 223.5183 *
      15) TX_RESP_Q145=NaO 2181   547971.9 232.7048 *
```

5.2.3 Árvore Temática 1 - Características do diretor

Além da visualização esquemática da árvore - que pode ser vista na Figura 12 - as variáveis envolvidas podem ser vistas na Figura - 13. Para a primeira árvore temática, em que as variáveis caracterizantes dos diretores das escolas são inseridas no modelo, nota-se a prevalência das das seguintes variáveis e seus respectivos efeitos dentro de seus nós:

- TX_RESP_Q001 - "Qual é a sua cor ou raça?"em que a localização de escolas com maior desempenho está à direita, nível da cor branca.

- TX_RESP_Q013 - "Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Atendimento aos pais ou responsáveis"aqui, para diretores que gastam mais de 3h semanais com essa atividade, há efeito positivo no desempenho.

- TX_RESP_Q006e7 -"Há quanto tempo você é diretor(a) desta escola? Anos"de forma geral, maiores períodos de experiência do diretor na escola afetam positivamente o desempenho.

- TX_RESP_Q017 - "Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Demandas da Secretaria de Educação"ao utilizar mais de 4h para essa atividade, há efeito positivo.

- TX_RESP_Q009 - "Quantas horas você trabalha em uma semana normal em atividades relacionadas à educação?"o trabalho por mais de 41h semanais, indicando possíveis horas extras, leva a melhores desempenhos.

- TX_RESP_Q020 - "Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Segurança"caso menos que 3h sejam dispendidas com a atividade, melhor o desempenho.

- TX_RESP_Q018 - "Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Merenda "caso menor que 3h sejam dispendidas com essa atividade, melhor o desempenho.

- TX_RESP_Q004e5 - "Você possui quanto tempo de experiência como diretor(a) de escola? Anos"de forma geral, maiores períodos que 9.9 anos de experiência como diretor afetam positivamente o desempenho.

Resumindo as condições, nota-se mais uma vez a questão racial dos diretores e também o efeito positivo de maiores tempos de experiência no cargo, seja na respectiva escola ou em outras instituições. Além disso, pode-se subentender que, quanto mais tempo é gasto com tarefas diretamente relacionadas ao cargo, como tratamento de questões com os responsáveis dos alunos e demandas da Secretaria da Educação, e menos tempo gasto com atividades adjacentes não necessariamente de sua tutela, como merenda e segurança escolar, melhor é o desempenho. Essas informações fazem-nos entender que uma gestão mais assertiva e com boa administração do tempo pelo diretor ou diretora são capazes de gerar efeito nos alunos e avaliações.

Figura 12: Árvore de Regressão: Árvore Temática 1

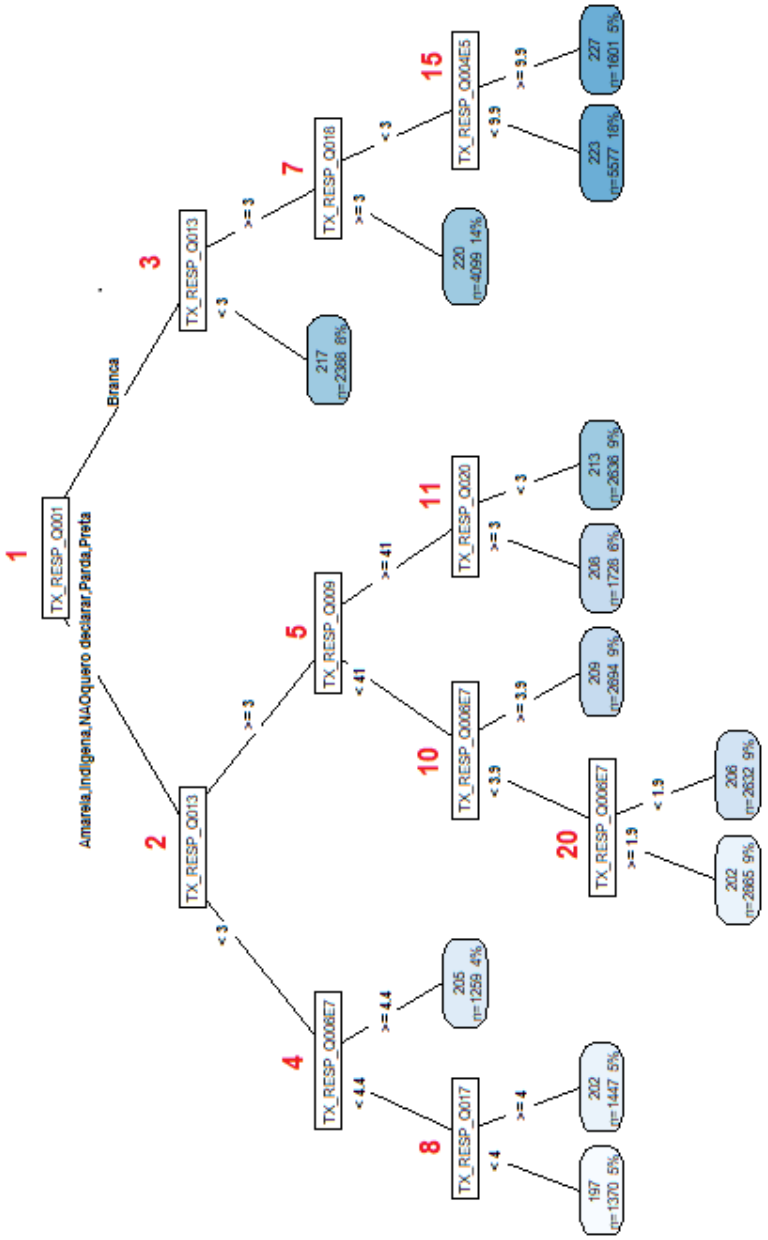


Figura 13: Print da Árvore Temática 1 em formato de texto

```
n= 30296

node), split, n, deviance, yval
      * denotes terminal node

 1) root 30296 18767980.0 212.9454
   2) TX_RESP_Q001=Amarela,Indigena,NAOquero declarar,Parda,Preta 16631 10712180.0 206.0454
     4) TX_RESP_Q013< 2.5 4076  2786778.0 201.5441
       8) TX_RESP_Q006E7< 4.375 2817  1879521.0 199.7808
        16) TX_RESP_Q017< 3.5 1370   896263.9 196.9628 *
        17) TX_RESP_Q017>=3.5 1447   962077.9 202.4488 *
       9) TX_RESP_Q006E7>=4.375 1259   878900.2 205.4896 *
     5) TX_RESP_Q013>=2.5 12555  7816002.0 207.5068
      10) TX_RESP_Q009< 40.5 8191  5392571.0 205.6728
        20) TX_RESP_Q006E7< 3.875 5497  3652675.0 204.1099
          40) TX_RESP_Q006E7>=1.875 2865  1916866.0 202.0438 *
          41) TX_RESP_Q006E7< 1.875 2632  1710265.0 206.3590 *
        21) TX_RESP_Q006E7>=3.875 2694  1699073.0 208.8618 *
      11) TX_RESP_Q009>=40.5 4364  2344170.0 210.9491
        22) TX_RESP_Q020>=2.5 1728   890760.1 207.6159 *
        23) TX_RESP_Q020< 2.5 2636  1421627.0 213.1341 *
   3) TX_RESP_Q001=Branca 13665  6300358.0 221.3430
     6) TX_RESP_Q013< 2.5 2388  1370095.0 216.9753 *
     7) TX_RESP_Q013>=2.5 11277  4875062.0 222.2679
      14) TX_RESP_Q018>=2.5 4099  1808804.0 219.6989 *
      15) TX_RESP_Q018< 2.5 7178  3023758.0 223.7349
        30) TX_RESP_Q004E5< 9.875 5577  2400783.0 222.7573 *
        31) TX_RESP_Q004E5>=9.875 1601   599076.3 227.1405 *
```

5.2.4 Árvore Temática 2 - Funcionamento e infraestrutura

Além da visualização esquemática da árvore - que pode ser vista na Figura 14 - as variáveis envolvidas podem ser vistas na Figura - 15. Para a segunda árvore temática, em que as variáveis sobre as condições de funcionamento e infraestrutura da escola são adicionadas no modelo, nota-se a prevalência das seguintes variáveis e seus respectivos efeitos dentro de seus nós:

- TX_RESP_Q097 - "Quais equipamentos existem nas áreas externas de recreação da sua escola? Escorregador"para a presença de escorregador, há efeito positivo no desempenho.

- TX_RESP_Q035 - "Indique quais são as etapas educacionais atendidas pela sua escola: Educação Infantil e Pré-escola (4 e 5 anos)."para o atendimento também de educação infantil na escola, há uma queda no desempenho.

- TX_RESP_Q080 - "Sobre a área externa da sua escola (pátio, parque e área verde), indique os itens existentes: Banheiro infantil"a existência desse tipo de banheiro relaciona-se com o maior desempenho.

- TX_RESP_Q083 - "Sobre a área externa da sua escola (pátio, parque e área verde), indique os itens existentes: Vegetação e jardim"a presença desses itens tem relação com o maior desempenho.

- TX_RESP_Q037 - "Indique quais são as etapas educacionais atendidas pela sua escola: Anos Finais do Ensino Fundamental."para o atendimento também do restante do ensino fundamental na escola, há uma queda no desempenho.

- TX_RESP_Q100 - "Quais equipamentos existem nas áreas externas de recreação da sua escola? Brinquedo para escalar"a presença desses itens tem relação com o maior desempenho.

- TX_RESP_Q101 - "Quais equipamentos existem nas áreas externas de recreação da sua escola? Banco"a presença desses itens tem relação com o maior desempenho.

- TX_RESP_Q087 - "Quais os tipos de solo ou revestimento do solo da área externa da sua escola? Cimento áspero"a utilização desse revestimento está relacionada a escolas com maior desempenho.

- TX_RESP_Q105 - "Avalie os seguintes aspectos da escola: O acesso à entrada principal das pessoas com deficiência física e visual (ex.: rampas e marcadores no chão)"a adequação desses aspectos, tem relação com aumento de desempenho.

- TX_RESP_Q045 - "Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: Os recursos pedagógicos foram suficientes."a concordância com a suficiência de recursos tem relação com um melhor desempenho.

Resumindo as condições, nota-se que escolas estão razoavelmente bem distribuídas para essas variáveis, de forma que o poder explicativo dessa seção não seja tão grande. De forma geral, a existência de alguns aspectos da infraestrutura, como banheiros, acesso facilitado e jardim, levam a desempenhos melhores, assim como o não atendimento de outras etapas educacionais na sede da escola.

Sobre isso, o melhor desempenho advindo da exclusividade de oferta de ensino fundamental em anos iniciais poderia ser interpretado pela possível dispensa de divisão de recursos e/ou profissionais entre as etapas da educação, dando maior enfoque e consequentemente um ensino de maior qualidade para os alunos desse tipo de escola. Por fim, a questão de recursos pedagógicos serem suficientes nas escolas também é um ponto importante, e sua falta pode desencadear pior desempenho.

Figura 14: Árvore de Regressão: Árvore Temática 2

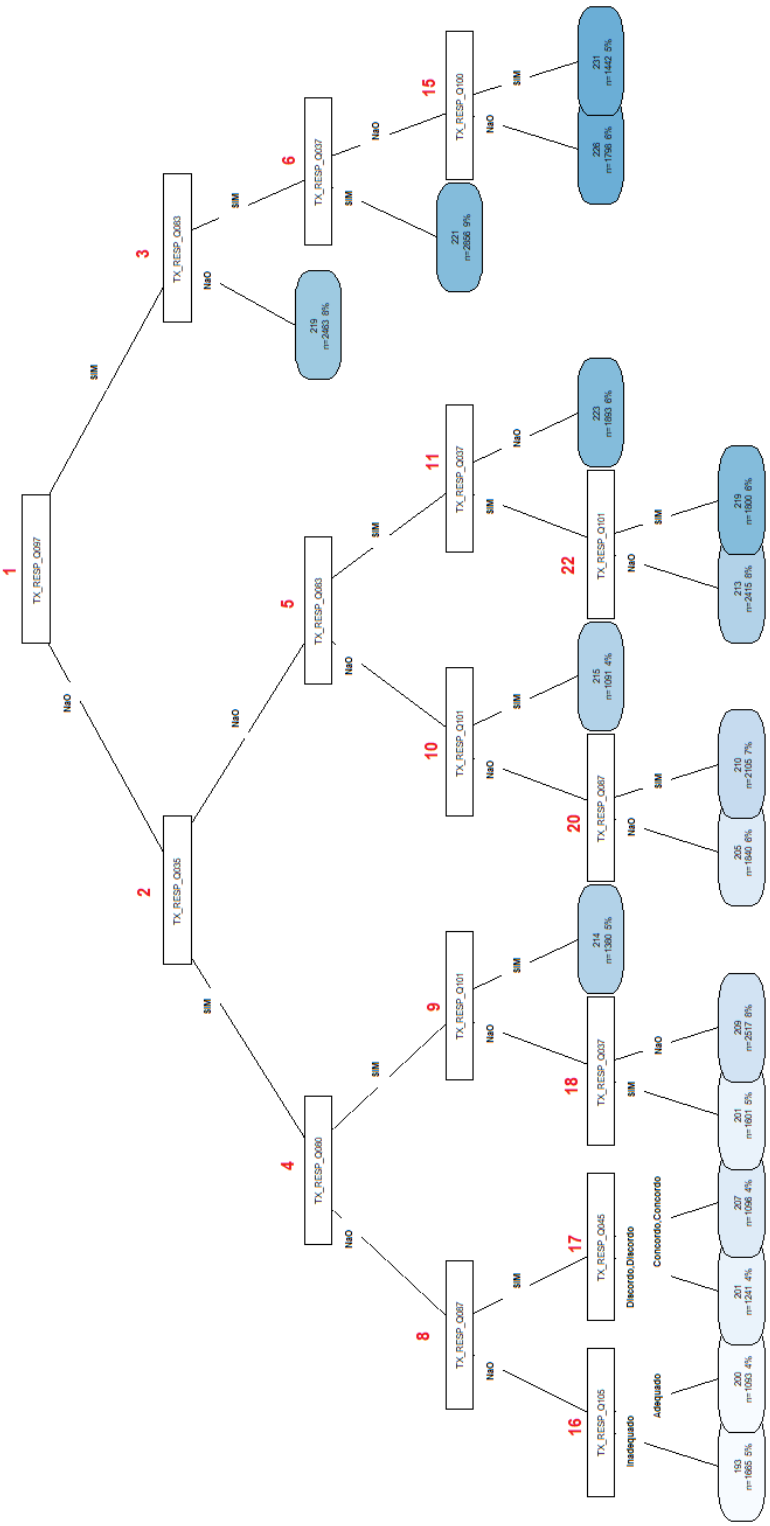


Figura 15: Print da Árvore Temática 2 em formato de texto

```
n= 30296

node), split, n, deviance, yval
      * denotes terminal node

 1) root 30296 18656580.0 213.0412
   2) TX_RESP_Q097=NaO 21737 13992710.0 209.1083
     4) TX_RESP_Q035=SIM 10593  7315592.0 203.8480
       8) TX_RESP_Q080=NaO 5095  3588003.0 199.5734
        16) TX_RESP_Q087=NaO 2758  1983166.0 196.0355
          32) TX_RESP_Q105=Inadequado 1665  1122260.0 193.2059 *
          33) TX_RESP_Q105=Adequado 1093   827267.6 200.3459 *
        17) TX_RESP_Q087=SIM 2337  1529575.0 203.7487
          34) TX_RESP_Q045=Discordo,Discordo  1241   724456.6 200.5497 *
          35) TX_RESP_Q045=Concordo,Concordo  1096   778037.4 207.3710 *
       9) TX_RESP_Q080=SIM 5498  3548222.0 207.8092
        18) TX_RESP_Q101=NaO 4118  2688940.0 205.7485
          36) TX_RESP_Q037=SIM 1601  1122492.0 200.8408 *
          37) TX_RESP_Q037=NaO 2517  1503360.0 208.8702 *
        19) TX_RESP_Q101=SIM 1380   789610.3 213.9586 *
     5) TX_RESP_Q035=NaO 11144  6105375.0 214.1085
      10) TX_RESP_Q083=NaO 5036  2838698.0 209.3344
        20) TX_RESP_Q101=NaO 3945  2209012.0 207.7299
          40) TX_RESP_Q087=NaO 1840  1074885.0 204.6612 *
          41) TX_RESP_Q087=SIM 2105  1101655.0 210.4123 *
        21) TX_RESP_Q101=SIM 1091   582805.5 215.1362 *
      11) TX_RESP_Q083=SIM 6108  3057261.0 218.0447
        22) TX_RESP_Q037=SIM 4215  2110475.0 216.0239
          44) TX_RESP_Q101=NaO 2415  1272773.0 213.4819 *
          45) TX_RESP_Q101=SIM 1800   801156.9 219.4346 *
        23) TX_RESP_Q037=NaO 1893   891249.8 222.5442 *
   3) TX_RESP_Q097=SIM 8559  3473749.0 223.0295
     6) TX_RESP_Q083=NaO 2463  1144428.0 218.5312 *
     7) TX_RESP_Q083=SIM 6096  2259345.0 224.8470
      14) TX_RESP_Q037=SIM 2856  1104406.0 221.1709 *
      15) TX_RESP_Q037=NaO 3240  1082323.0 228.0874
        30) TX_RESP_Q100=NaO 1798   665709.3 225.9376 *
        31) TX_RESP_Q100=SIM 1442   397942.5 230.7680 *
```

5.2.5 Árvore Temática 3 - Gestão e inclusão

Além da visualização esquemática da árvore - que pode ser vista na Figura 16 - as variáveis envolvidas podem ser vistas na Figura - 17. Para a terceira árvore temática, em que as variáveis sobre a gestão pedagógica e educação inclusiva na escola são inseridas no modelo, nota-se a prevalência das seguintes variáveis e seus respectivos efeitos dentro de seus nós:

- TX_RESP_Q184 - "Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Tempo de serviço."no caso da inutilização desse critério, o desempenho varia para notas menores.

- TX_RESP_Q153 - "Todos(as) conseguem se alimentar sentados"no caso em que há discordância dessa constatação, o desempenho também tem efeito negativo.

- TX_RESP_Q166 - "Neste ano e nesta escola, todos que solicitaram vagas conseguiram se matricular?"Apesar de pouco intuitivo, quando a escola não pode matricular a todos solicitantes, seu desempenho tem efeito positivo.

- TX_RESP_Q202 - "Neste ano, para a redução da repetência escolar, avalie o resultado das seguintes ações realizadas nesta escola: Oferta de reforço escolar"A oferta de reforço nas escolas tem impacto positivo no desempenho dessas.

- TX_RESP_Q145 - "Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Contribuições voluntárias dos(as) profissionais da escola."Quando ocorre esse tipo de contribuição de recursos o desempenho da escola é afetado de forma negativa.

- TX_RESP_Q222- "Nesta escola, há projetos com as seguintes temáticas: Educação financeira e consumo sustentável"A existência de projetos com essa temática tem relação com um melhor desempenho das escolas.

- TX_RESP_Q118 - "O Conselho Escolar é um colegiado geralmente constituído por representantes da escola e da comunidade que tem como objetivo acompanhar as atividades escolares. Na sua escola existe Conselho Escolar?"A existência de um conselho escolar ativo tem relação com o maior desempenho escolar.

- TX_RESP_Q147 - "Quantas refeições são oferecidas nesta escola para alunos(as) que permanecem menos de 4 horas na escola"Maior a quantidade de refeições oferecida (diferente de uma única refeição) temos um melhor desempenho.

- TX_RESP_Q210 - "Nesta escola, há projetos com as seguintes temáticas: Sexualidade"A existência de projetos com essa temática está relacionado a um desempenho mais alto.

- TX_RESP_Q186 - "Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Professores(as) experientes nas turmas com facilidade de aprendizagem."Quando esse critério não é utilizado, o desempenho varia para valores mais altos de forma geral.

- TX_RESP_Q188 - "Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Manutenção do(a) professor(a) com a mesma turma."O uso desse critério para alocação de turmas reflete em ligeiramente menores desempenhos.

- TX_RESP_Q144 -"Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Contribuições voluntárias dos familiares dos(as) estudantes."O fornecimento de recursos de forma voluntária está relacionado a um ligeiro melhor desempenho.

- TX_RESP_Q122 - "Considere os atores relacionados a seguir e indique quantos participam do Conselho Escolar: Pais (ou responsáveis)"Para maiores valores de participantes do conselho escolar, há um melhor desempenho.

Resumindo as condições, nota-se que escolas estão razoavelmente bem distribuídas para essas variáveis. De forma geral, a análise das variáveis ligadas à gestão pedagógica e à educação inclusiva revela que determinados fatores, como a oferta de reforço escolar, projetos sobre educação financeira e sexualidade, além da existência de um conselho escolar ativo, estão associados a melhores desempenhos. Esses elementos podem indicar uma gestão mais estruturada e um ambiente educacional mais abrangente e integrado. Por outro lado, critérios como a atribuição de professores com base no tempo de serviço ou a manutenção de um professor com a mesma turma não necessariamente favorecem o desempenho, o que sugere que a rigidez na distribuição de turmas pode limitar a flexibilidade necessária para otimizar o processo educacional.

A contribuição voluntária de recursos por parte de profissionais, embora pareça indicar envolvimento, está associada a um desempenho inferior, enquanto a participação dos pais no conselho escolar e a oferta de mais refeições estão ligadas a um melhor rendimento, talvez refletindo uma maior integração entre a escola e a comunidade ou ainda o suprimento de necessidades dos alunos.

Figura 16: Árvore de Regressão: Árvore Temática 3

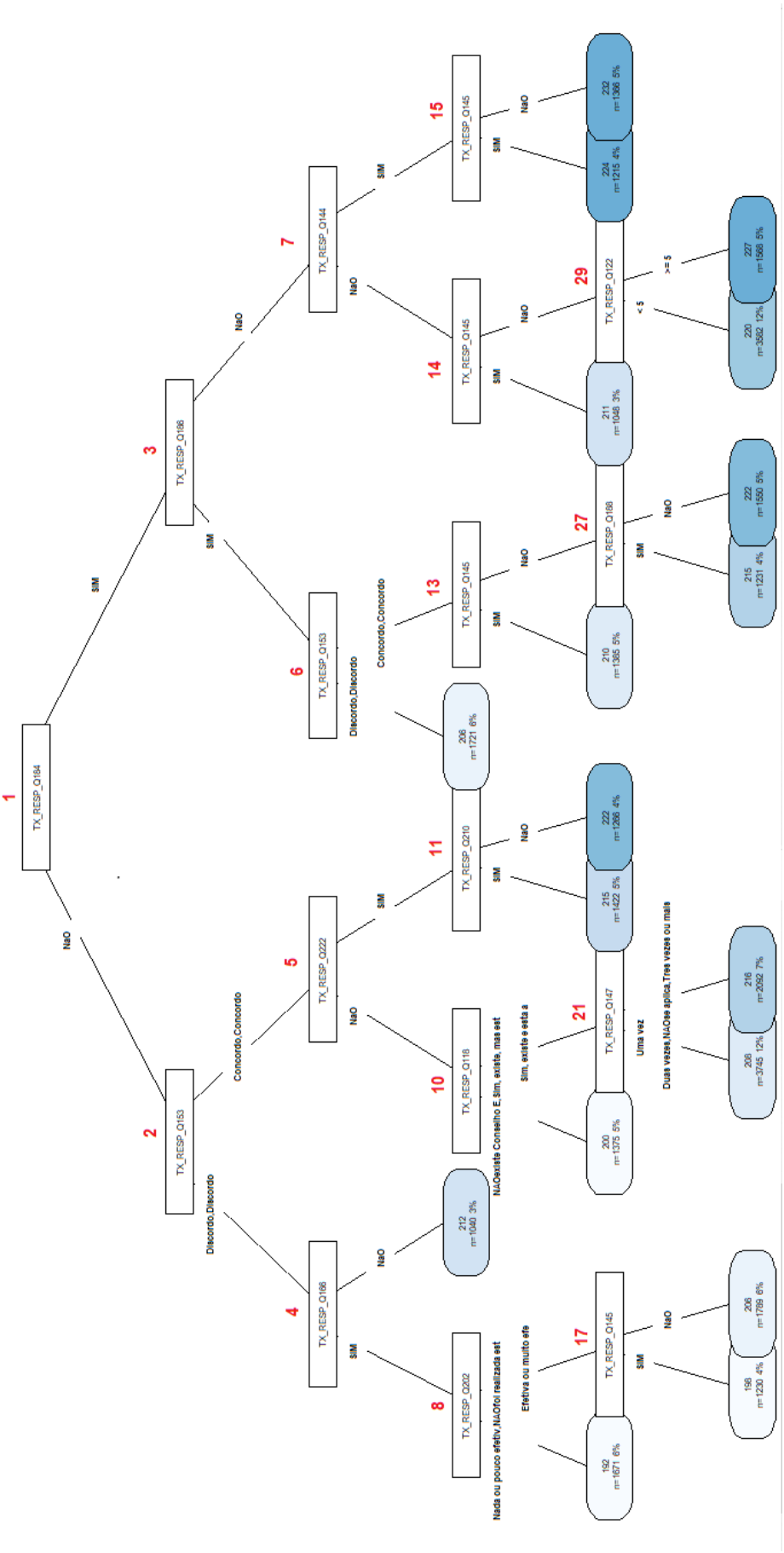
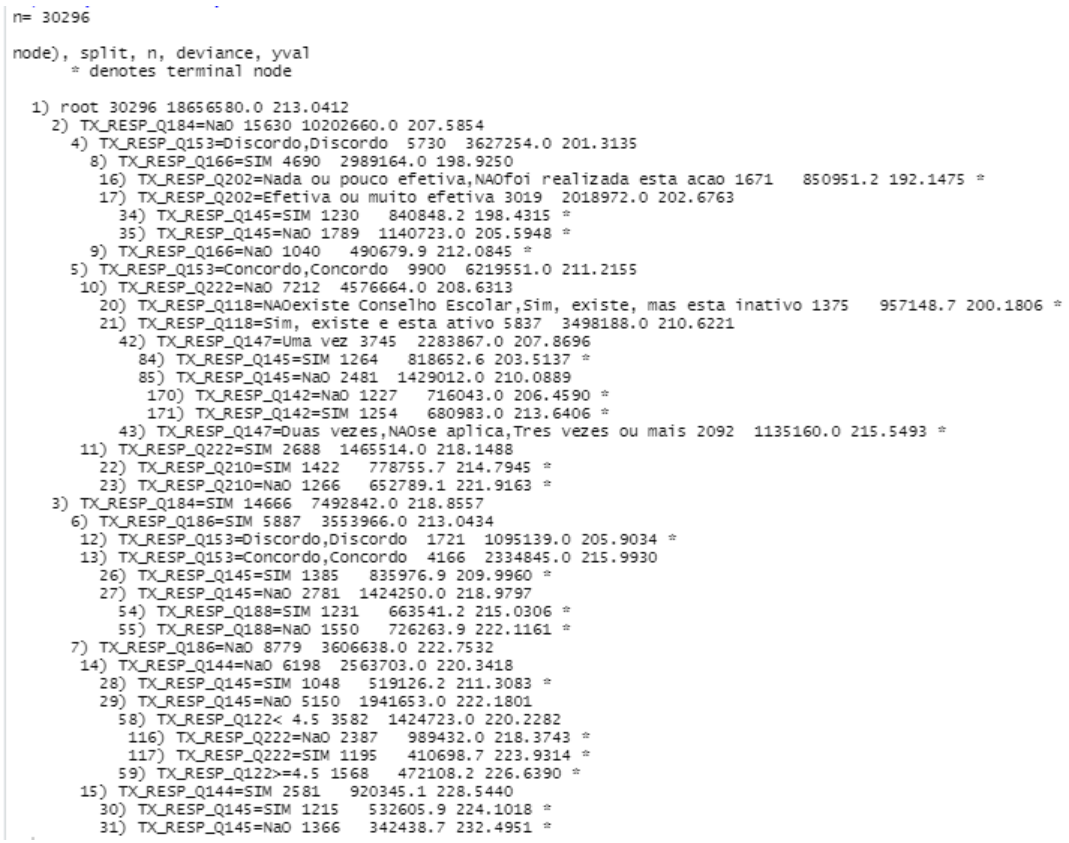


Figura 17: Print da Árvore Temática 3 em formato de texto



```
n= 30296

node), split, n, deviance, yval
      * denotes terminal node

  1) root 30296 18656580.0 213.0412
    2) TX_RESP_Q184=NaO 15630 10202660.0 207.5854
      4) TX_RESP_Q153=Discordo,Discordo  5730  3627254.0 201.3135
        8) TX_RESP_Q166=SIM 4690  2989164.0 198.9250
         16) TX_RESP_Q202=Nada ou pouco efetiva,NAOfoi realizada esta acao 1671   850951.2 192.1475 *
         17) TX_RESP_Q202=Efetiva ou muito efetiva 3019  2018972.0 202.6763
           34) TX_RESP_Q145=SIM 1230   840848.2 198.4315 *
           35) TX_RESP_Q145=NaO 1789  1140723.0 205.5948 *
        9) TX_RESP_Q166=NaO 1040   490679.9 212.0845 *
      5) TX_RESP_Q153=Concordo,Concordo  9900  6219551.0 211.2155
       10) TX_RESP_Q222=NaO 7212  4576664.0 208.6313
         20) TX_RESP_Q118=NAOexiste Conselho Escolar,Sim, existe, mas esta inativo 1375   957148.7 200.1806 *
         21) TX_RESP_Q118=Sim, existe e esta ativo 5837  3498188.0 210.6221
           42) TX_RESP_Q147=Uma vez 3745  2283867.0 207.8696
             84) TX_RESP_Q145=SIM 1264   818652.6 203.5137 *
             85) TX_RESP_Q145=NaO 2481  1429012.0 210.0889
              170) TX_RESP_Q142=NaO 1227   716043.0 206.4590 *
              171) TX_RESP_Q142=SIM 1254   680983.0 213.6406 *
           43) TX_RESP_Q147=Duas vezes,NAOse aplica,Tres vezes ou mais 2092  1135160.0 215.5493 *
       11) TX_RESP_Q222=SIM 2688  1465514.0 218.1488
         22) TX_RESP_Q210=SIM 1422   778755.7 214.7945 *
         23) TX_RESP_Q210=NaO 1266   652789.1 221.9163 *
    3) TX_RESP_Q184=SIM 14666  7492842.0 218.8557
      6) TX_RESP_Q186=SIM 5887  3553966.0 213.0434
       12) TX_RESP_Q153=Discordo,Discordo  1721  1095139.0 205.9034 *
       13) TX_RESP_Q153=Concordo,Concordo  4166  2334845.0 215.9930
         26) TX_RESP_Q145=SIM 1385   835976.9 209.9960 *
         27) TX_RESP_Q145=NaO 2781  1424250.0 218.9797
           54) TX_RESP_Q188=SIM 1231   663541.2 215.0306 *
           55) TX_RESP_Q188=NaO 1550   726263.9 222.1161 *
      7) TX_RESP_Q186=NaO 8779  3606638.0 222.7532
       14) TX_RESP_Q144=NaO 6198  2563703.0 220.3418
         28) TX_RESP_Q145=SIM 1048   519126.2 211.3083 *
         29) TX_RESP_Q145=NaO 5150  1941653.0 222.1801
           58) TX_RESP_Q122< 4.5 3582  1424723.0 220.2282
            116) TX_RESP_Q222=NaO 2387   989432.0 218.3743 *
            117) TX_RESP_Q222=SIM 1195   410698.7 223.9314 *
           59) TX_RESP_Q122>=4.5 1568   472108.2 226.6390 *
       15) TX_RESP_Q144=SIM 2581   920345.1 228.5440
         30) TX_RESP_Q145=SIM 1215   532605.9 224.1018 *
         31) TX_RESP_Q145=NaO 1366   342438.7 232.4951 *
```

5.3 Avaliação de precisão

Para as árvores modeladas, foram também desenvolvidas formas de avaliar sua precisão e complexidade. Inicialmente, foram construídos e plotados os Gráficos de Custo de Complexidade da árvore em função do erro relativo dos seus valores. Retomando o conceito, o parâmetro de complexidade (CP) é definido como o valor mínimo pelo qual a redução de erro deve ser alcançada para que uma divisão seja mantida na árvore, CP representa uma penalidade para cada terminal adicionado na árvor. A complexidade aqui é referente à quantidade de nós terminais da árvore - *size of tree* - e caracteriza uma espécie de *trade-off* entre o tamanho da árvore e o quanto mais nós são capazes de diminuir o erro entre os valores efetivos e os previstos.

Os gráficos exibem o erro relativo da árvore em função do valor de CP, com o eixo Y representando o erro relativo e o eixo X representando o valor de CP. Cada ponto no gráfico representa uma subárvore com um certo número de nós terminais. A subárvore correspondente ao menor erro relativo é geralmente considerada a melhor árvore, mas pontos onde o erro já não reduz significativamente com a diminuição de CP ou encontra-

se estabilizado também são boas opções. Esses gráficos gerados apresentam-se a seguir, nas Figuras 18, 19, 20, 21 e 22.

Figura 18: Gráfico Custo Complexidade x Erro Relativo: Árvore Geral

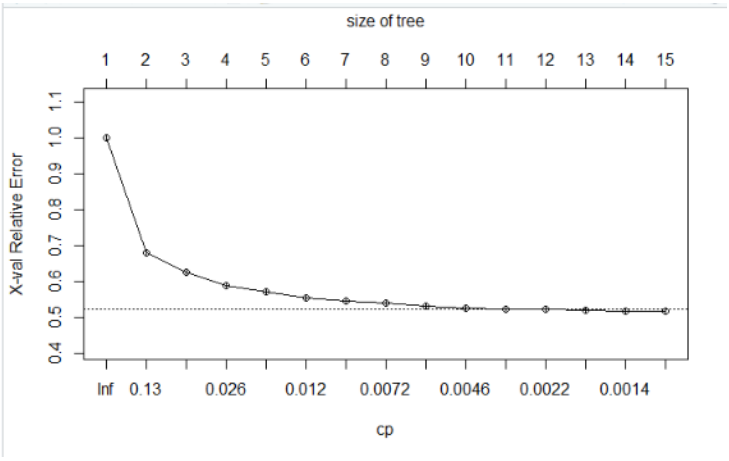


Figura 19: Gráfico Custo Complexidade x Erro Relativo: Árvore Geral Questionário Diretor

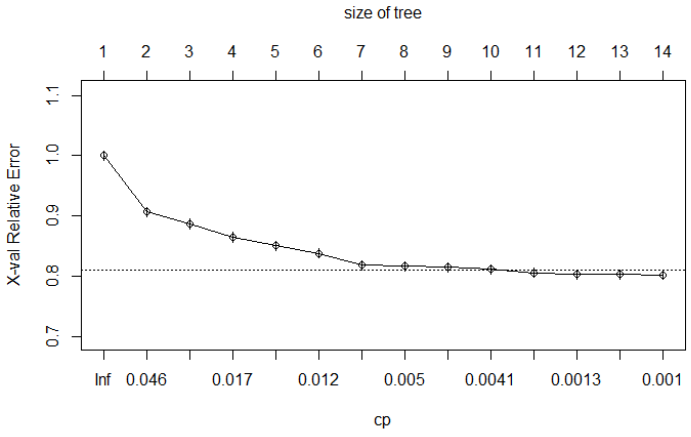


Figura 20: Gráfico Custo Complexidade x Erro Relativo: Árvore Temática 1

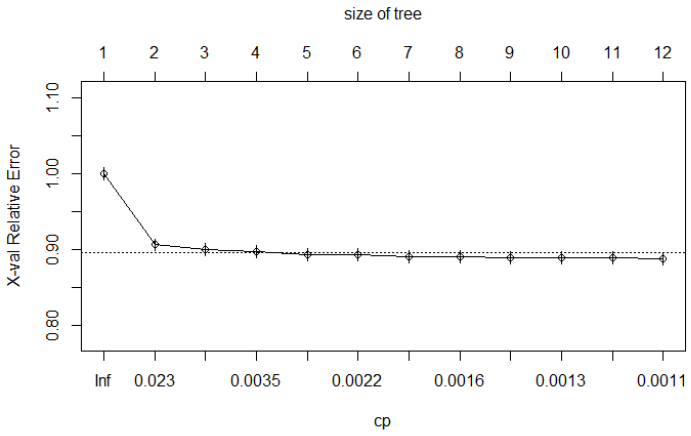


Figura 21: Gráfico Custo Complexidade x Erro Relativo: Árvore Temática 2

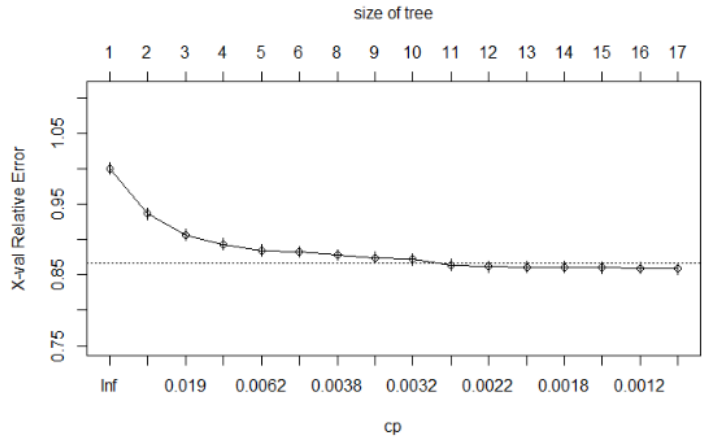
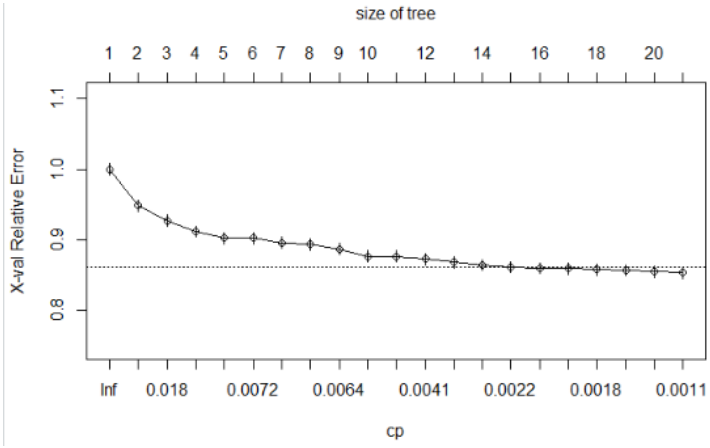


Figura 22: Gráfico Custo Complexidade x Erro Relativo: Árvore Temática 3



Esses gráficos apontam ainda que o valor estipulado como parâmetro no *rpart* para o cp = 0.001 abrange a estabilização do erro de modo que não perde-se melhorias de desempenho do modelo por conta de tal parâmetro. Por fim, sobre esse parâmetro, entende-se que a Árvores Geral é capaz de gerar erros menores (cerca de 0.5) que as demais árvores. As árvores Geral Questionário Diretor e Temáticas 1 (Caracterização do diretor), 2 (Infraestrutura) e 3 (Gestão pedagógica) apresentam erros maiores, variando entre 0.80 e 0.90 aproximadamente. Isso indica que não há um poder preditivo alto para todos os casos, mas que existem conjuntos de variáveis mais bem relacionados com o desempenho escolar.

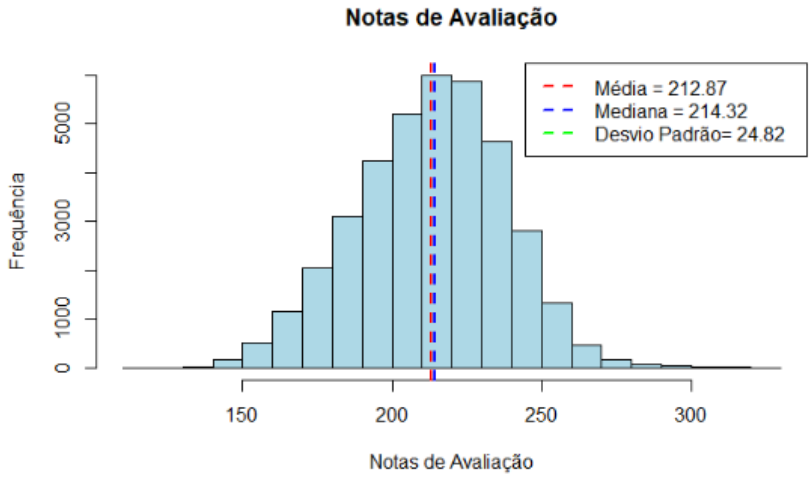
Além disso, para cada modelo também foram calculados os Erro Médio Absoluto (MAE), Raiz do Erro Quadrático Médio (RMSE) e o coeficiente de determinação (R^2), como indicado na Tabela 32 - Avaliação de Precisão:

Tabela 32: Avaliação de Precisão

Árvore	MAE	RMSE	R^2
Geral	13.61	17.71	0.49
Geral Questionário Diretor	17.31	22.02	0.19
Temática 1	18.25	23.08	0.12
Temática 2	18.12	22.88	0.15
Temática 3	18.06	22.86	0.15

Também foi criada a visualização da distribuição das notas reais da avaliação, a qual encontra-se na Figura 23 - Histograma notas unificadas 5º ano EF. Nela, entende-se que as notas possuem média de 212.87 pontos e um desvio padrão de 24.82. O valor máximo para a nota das escolas na avaliação foi de 321.75 e o mínimo de 118.96.

Figura 23: Histograma notas unificadas 5º ano EF



Dessa forma, tendo os valores apresentados como base, é possível verificar que:

- O **MAE** mede a média dos erros absolutos entre as previsões do modelo e os valores reais. Quanto menor o valor, melhor a precisão do modelo, pois significa que os erros médios são menores. O MAE para a Árvore Geral é de 13.61 o que indica que, em termos de erro médio absoluto, esse modelo tem um desempenho melhor quando comparado aos demais. As árvores Diretor e Temáticas 1, 2 e 3 apresentam um MAE mais alto (17.31, 18.25, 18.12 e 18.06), sugerindo que essas árvores têm uma precisão inferior, com erros médios maiores.

- O **RMSE** mede a raiz quadrada da média dos erros quadráticos. Assim como o

MAE, valores menores indicam maior precisão, mas o RMSE penaliza mais os erros grandes. Apesar do modelo da Árvore Geral apresentar um erro menor, os demais modelos gerados não apresentam a variabilidade de suas previsões muito melhorada quando comparada a dos dados reais observados, tendo em vista que os valores do RMSE são razoavelmente próximos do desvio padrão dos dados (cerca de 22 para o RMSE *versus* 24.8 para o desvio padrão). Em outras palavras, as previsões dos modelos têm uma variabilidade de erro que é comparável à variabilidade natural dos dados reais - o que é um sinal de baixo desempenho preditivo.

- O R^2 mede o quão bem o modelo consegue explicar a variabilidade dos dados. Valores mais próximos de 1 indicam que o modelo explica a maior parte da variação nos dados. A Árvore Geral tem um R^2 de 0.49, o que significa que esse modelo explica cerca de 49% da variabilidade dos dados. Embora não seja perfeito, ainda é uma explicação razoável. Já as demais têm um R^2 muito menor (de 0.19 a 0.12), sugerindo que esses modelos explicam menos que 20% da variabilidade, o que é considerado um desempenho fraco. Isso significa que grande parte da variação nos dados não está sendo explicada por essas árvores.

Em resumo, a Árvore Geral possui um desempenho moderado, dado a presença das variáveis de localização no modelo, as quais se mostraram mais relevantes que as demais. Já as outras árvores têm maiores erros (MAE e RMSE) e explicam muito menos a variabilidade dos dados (R^2), mostrando que esses modelos são menos eficazes em prever o desempenho - efeito também da retirada das variáveis demográficas. Com isso, reforça-se que os modelos aqui desenvolvidos cumprem com o objetivo de evidenciar a influência de certas variáveis e não necessariamente são bons para a previsão das notas das escolas com alta precisão de valores, dado que não são capazes de gerar previsões com muito menos variabilidade de erro que os dados reais.

5.4 Distribuição geográfica dos melhores desempenhos

Para melhor visualizar a localização das escolas de maiores desempenhos previstos, retomou-se a região das escolas em que estão os melhores desempenhos conforme os modelos elaborados. Para a obtenção destas tabelas, foi necessários filtrar os dados da base geral seguindo o caminho da respectiva árvore que levasse ao maior desempe-

nho previsto. Com isso, tem-se as Tabelas 33, 34,35, 36 e 37 e caminhos (do nó raiz ao terminal):

- **ÁRVORE GERAL** - Nota 241

  ID_UF: AC, CE, DF, ES, GO, MG, MS, PR, RJ, RS, SC, SP

  NIVEL_SOCIO_ECONOMICO: NÍVEIS VI, VII

Tabela 33: Localização das melhores escolas - Árvore Geral

Região	Contagem de Escolas	%
Centro-Oeste	71	3%
Nordeste	1	0%
Norte	1	0%
Sudeste	937	40%
Sul	1328	57%
Total Geral	2338	100%

- **ÁRVORE GERAL QUESTIONÁRIO DIRETOR** - Nota 233

  TX_RESP_Q001 - Qual é a sua cor ou raça: BRANCA

  TX_RESP_Q144 - Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Contribuições voluntárias dos familiares dos(as) estudantes: SIM

  TX_RESP_Q145 - Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Contribuições voluntárias dos(as) profissionais da escola: NÃO

Tabela 34: Localização das melhores escolas - Árvore Geral Questionário Diretor

Região	Contagem de Escolas	%
Centro-Oeste	47	2%
Nordeste	27	1%
Norte	16	1%
Sudeste	1124	42%
Sul	1455	55%
Total Geral	2669	100%

- **ÁRVORE TEMÁTICA 1** - Nota 227

  TX_RESP_Q001 - Qual é a sua cor ou raça: BRANCA

  TX_RESP_Q013 - Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Atendimento aos pais ou responsáveis: MAIOR OU IGUAL A 3 HORAS

  TX_RESP_Q018 - Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Merenda: MAIOR OU IGUAL A 3 HORAS

  TX_RESP_Q004E5 - Você possui quanto tempo de experiência como diretor(a) de escola: MAIOR OU IGUAL A 9.9 ANOS

Tabela 35: Localização das melhores escolas - Árvore Temática 1

Região	Contagem de Escolas	%
Centro-Oeste	65	3%
Nordeste	236	12%
Norte	65	3%
Sudeste	1145	59%
Sul	432	22%
Total Geral	1943	100%

- **ÁRVORE TEMÁTICA 2** - Nota 231

  TX_RESP_Q097 - Quais equipamentos existem nas áreas externas de recreação da sua escola? Escorregador: SIM

  TX_RESP_Q083 - Sobre a área externa da sua escola (pátio, parque e área verde), indique os itens existentes: Vegetação e jardim: SIM

  TX_RESP_Q037 - Indique quais são as etapas educacionais atendidas pela sua escola: Anos Finais do Ensino Fundamental: NÃO

  TX_RESP_Q100 - Quais equipamentos existem nas áreas externas de recreação da sua escola? Brinquedo para escalar: SIM

Tabela 36: Localização das melhores escolas - Árvore Temática 2

Região	Contagem de Escolas	%
Centro-Oeste	122	9%
Nordeste	48	3%
Norte	33	2%
Sudeste	706	50%
Sul	517	36%
Total Geral	1426	100%

- **ÁRVORE TEMÁTICA 3** - Nota 232

  TX_RESP_Q184 - Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Tempo de serviço: SIM

  TX_RESP_Q186 - Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Professores(as) experientes nas turmas com facilidade de aprendizagem: NÃO

  TX_RESP_Q144 - Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Contribuições voluntárias dos familiares dos(as) estudantes: SIM

  TX_RESP_Q145 - Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Contribuições voluntárias dos(as) profissionais da escola: NÃO

Tabela 37: Localização das melhores escolas - Árvore Temática 3

Região	Contagem de Escolas	%
Centro-Oeste	49	3%
Nordeste	8	1%
Norte	6	0%
Sudeste	771	49%
Sul	745	47%
Total Geral	1579	100%

Com essas tabelas, entende-se que para todos os caminhos de melhor desempenho, as escolas "modelo"concentram-se principalmente nas regiões Sudeste e Sul. Especificamente nos estados:

- Para a Árvore Geral: São Paulo - com 28% de todas as escolas de bom desempenho - e Rio Grande do Sul com 22%
- Para a Árvore Geral Questionário Diretor: São Paulo (38%) e Santa Catarina (20%)
- Para a Árvore Temática 1 (diretor): São Paulo (47%) e Minas Gerais (10%)
- Para a Árvore Temática 2 (infraestrutura): São Paulo (35%) e Paraná (27%)
- Para a Árvore Temática 3 (gestão): São Paulo (43%) e Santa Catarina (21%)

Com essa distribuição, nota-se que o estado de São Paulo poderia ser utilizado para demarcar um modelo de bom desempenho (enquanto apresenta previsões mais altas que a média geral das notas). Além disso, é possível verificar que, dessas escolas de SP, cerca de 95% delas encontram-se em área urbana para todas as árvores. Dessa forma, é possível elencá-las como escolas cujas características poderiam ser reproduzidas em outros estados ou situações de baixo desempenho. No entanto, somente com as informações da SAEB não é possível aprofundar a definição das escolas modelo ao nível de rua, bairro e nome da escola, pois esses dados de localização são mascarados nas bases públicas utilizadas.

5.5 *Insights* gerais

Esse tópico foi criado para compilar as principais informações sobre as árvores de regressão geradas e a hierarquia das variáveis envolvidas em cada árvore.

A análise da árvore de regressão geral revelou que as variáveis **demográficas das escolas, como a localização estadual (ID_UF), nível socioeconômico, região e tipo de localização (rural ou urbano)**, são predominantes na explicação da variação de desempenho escolar. Escolas nas regiões Norte, Nordeste e Centro-Oeste, com nível socioeconômico mais baixo, tendem a apresentar resultados inferiores.

Na árvore geral do questionário diretor, em que todas as variáveis com exceção daquelas de localização da escola foram utilizadas, observou-se que fatores como **a presença de um projeto político-pedagógico, grêmio estudantil, brinquedos de recreação e áreas verdes como jardim ou grama se associam a um melhor desempenho escolar**, o que reflete uma estrutura escolar mais preparada e equipada. Contribuições

financeiras dos responsáveis dos alunos mostraram-se positivas para o desempenho, enquanto o apoio financeiro dos próprios funcionários teve efeito negativo. Além disso, há um fator relevante na **cor ou raça do diretor**, onde diretores brancos aparentam ter efeito positivo; no entanto, sugere-se que esse efeito reflete desigualdades sociais e não uma relação causal, merecendo análise mais profunda dos contextos socioeconômicos das escolas.

Na primeira árvore temática, que inclui variáveis demográficas do diretor, destacam-se fatores como a raça do diretor, **maior experiência no cargo, e a distribuição de tempo entre atividade**s. Atividades diretamente ligadas ao cargo, quando recebem maior atenção, também contribuem para melhor desempenho. Em contraste, menos tempo dedicado a tarefas menos centrais, como segurança e merenda, está associado a melhores resultados, sugerindo que uma **gestão mais focada e eficiente** beneficia os alunos e o desempenho geral da escola.

Na segunda árvore temática, que aborda as condições de funcionamento e infraestrutura, variáveis como **a presença de equipamentos recreativos (escorregadores, brinquedos para escalar), áreas externas com vegetação, banheiros infantis e acessibilidade** destacaram-se positivamente. Por outro lado, sugere-se que escolas focadas exclusivamente nos anos iniciais do ensino fundamental podem ter maior eficiência ao não dividir recursos entre diferentes etapas. A **adequação pedagógica e a disponibilidade de recursos** também aparecem como fatores essenciais para um melhor desempenho.

Por fim, a terceira árvore temática, que explora a gestão pedagógica e educação inclusiva, indicou que variáveis como a **oferta de reforço escolar, projetos sobre educação financeira e sexualidade, e a existência de um conselho escolar ativo** influenciam positivamente o desempenho. **A participação ativa dos pais no conselho escolar e a oferta de refeições adicionais** foram associadas a melhores resultados, possivelmente refletindo maior integração entre escola, comunidade e o suprimento de necessidades básicas dos alunos.

Além disso, a análise da distribuição geográfica das escolas com **melhor desempenho**, baseada nas árvores de decisão geradas, indicou uma **concentração significativa dessas escolas nas regiões Sudeste e Sul do Brasil**, especialmente nos estados de São Paulo e Santa Catarina, que apresentam alta proporção de escolas de bom desem-

penho. **São Paulo se destaca como modelo**, sugerindo que as características dessas escolas poderiam ser replicadas em outras regiões para melhorar o desempenho nacional. Contudo, a ausência de dados para a identificação exata dessas escolas limita uma análise mais detalhada.

Com esses pontos, configurou-se uma análise baseada nas informações, em grande parte qualitativas, dadas nos formulários. Essa análise inicial forneceu informações valiosas, que ajudaram a direcionar o foco para variáveis específicas - as quais poderiam ser investigadas de forma mais aprofundada, com intuito de validar com justificativas robustas as hipóteses de causa e consequência geradas aqui.

Identificou-se ainda a importância de explorar com maior detalhamento a frequência de certos eventos e o modo como determinados processos ocorrem na prática. Isso sugere que uma análise qualitativa de algumas variáveis entendidas como categóricas no formulário ou um destrinchamento maior de contexto poderia melhorar a capacidade preditiva, ao oferecer uma visão mais precisa sobre as relações entre as variáveis.

Por fim, a análise dos fatores impactantes no desempenho escolar indica a possibilidade de recomendações práticas para gestores educacionais brasileiros. Com base nas variáveis identificadas, sugere-se que investimentos sejam direcionados para melhorar a infraestrutura escolar, especialmente em equipamentos recreativos e áreas verdes, que contribuem positivamente para o desempenho dos alunos. Além disso, a promoção de uma maior integração entre a escola e a comunidade, por meio de conselhos escolares ativos e da participação dos pais, poderia fortalecer a relação entre os agentes educacionais e o ambiente familiar, refletindo em melhores resultados acadêmicos. Para escolas em regiões com nível socioeconômico mais baixo, políticas de apoio específicas (como a criação de projetos pedagógicos voltados para reforço escolar) podem auxiliar no enfrentamento das desigualdades regionais. Entende-se ainda que a implementação dessas estratégias no cenário educacional brasileiro requer um planejamento cuidadoso, adaptado às diferentes realidades escolares, e o compartilhamento de boas práticas de gestão.

6 CONCLUSÕES

O trabalho desenvolvido buscou responder à questão "O que tem influenciado o desempenho brasileiro na educação?" por meio de uma abordagem de ciência de dados aplicada a grandes volumes de informações sobre as escolas brasileiras. Assim, foi proposto um método de análise de dados, baseado na utilização de Árvores de Regressão, para identificar variáveis que explicassem a variabilidade no desempenho escolar e evidenciar fatores institucionais além das capacidades individuais dos alunos - como infraestrutura, gestão de recursos e aspectos geográficos das escolas brasileiras - capazes de impactar positivamente ou negativamente as notas do SAEB, a fonte de dados utilizada.

O método de análise escolhido se mostrou adequado frente à dimensão do conjunto de dados, tendo em vista as limitações no uso eficaz de métodos estatísticos e/ou de regressão convencionais em bases de dados com tantas variáveis e níveis envolvidos. Também notou-se que técnicas mais robustas, como de aprendizado de máquina, embora fossem valiosas para alcançar resultados mais precisos, possivelmente sacrificariam a interpretabilidade dos resultados, algo essencial para o estudo. Desta forma, a escolha pelas Árvores de Regressão equilibrou esses aspectos e ofereceu um modelo replicável para outros conjuntos de dados e ajustável a diferentes contextos. Esse equilíbrio permite um modelo analítico útil tanto para cientistas de dados em outros estudos de caso quanto para formuladores de políticas públicas.

Os resultados obtidos indicaram características-chave para escolas com melhor desempenho, como qualidade da infraestrutura, suporte pedagógico e gestão ativa de recursos - revisitar a Seção 5.5 - *Insights* gerais para mais detalhes. Vale ressaltar que não se buscava com esse trabalho uma solução absoluta para as lacunas educacionais do país, mas, ao definir perfis e características de escolas modelo seria possível realizar um direcionamento mais eficiente de esforços e recursos para melhorar os índices de desempenho. O trabalho também evidenciou a importância de explorar tanto os aspectos qualitativos nos formulários utilizados quanto os qualitativos, e sugere que uma análise de variáveis mais detalhadas poderia aumentar a precisão dos modelos e oferecer elucidações ainda mais robustas sobre desempenho escolar.

Vale ressaltar também que, apesar de as conclusões e *insights* obtidos por meio das árvores de regressão parecerem, à primeira vista, intuitivos ou até mesmo óbvios, a capacidade de embasá-los e comprová-los com dados rigorosos representa um benefício

significativo para a formulação de políticas públicas efetivas. A análise evidencia que o impacto positivo de "fazer o básico bem feito"— como garantir uma infraestrutura escolar adequada, uma gestão ativa e suporte pedagógico consistente — pode ser mensurado e comprovado. Esse suporte analítico permite que decisões sejam fundamentadas em dados concretos e evidências através de um método talvez pouco trivial para elaboração de políticas.

O desenvolvimento deste trabalho também proporcionou uma ampliação do conhecimento da aluna sobre tratamento, modelagem e análise de dados aplicada à educação e possibilitou uma reflexão crítica sobre as limitações e potenciais melhorias no uso dessas informações. A associação de um estudo de caso para a aplicação prática dos conhecimentos foi de suma importância na absorção dos conteúdos, dado que apenas a realização de uma revisão teórica não teria permitido esse nível de detalhamento. O trabalho contribui para a engenharia de produção ao demonstrar como melhorar sistemas complexos e apoiar decisões baseadas em dados. A aplicação prática das técnicas estudadas neste trabalho de formatura sublinha o papel transformador que a ciência de dados pode desempenhar na solução de problemas também de cunho social.

Por fim, como continuidade deste estudo, sugere-se aprofundar a análise com coleta de informações mais acionáveis dos dados da educação (como informações quantitativas de recursos faltantes, posicionamento específico da escola, entre outros) e expandir o modelo para outras bases abrangentes, o que permite a aplicação do método de análise de dados proposto em variados projetos de políticas públicas ou ainda em diferentes contextos envolvendo *Big Data*.

REFERÊNCIAS*

1 MCAFEE, A.; BRYNJOLFSSON, E. Big data: The management revolution. Harvard Business Review, v. 90, n. 10, p. 60–68, 2012. Disponível em: ⟨https://hbr.org/2012/10/big-data-the-management-revolution⟩.

2 MANYIKA, J. et al. Big data: The next frontier for innovation, competition, and productivity. 2011. Disponível em: ⟨https://www.researchgate.net/publication/271727114_What_is_Big_Data_and_Why_is_It_Important⟩.

3 Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Resultados do Programa Internacional de Avaliação de Estudantes de 2022. 2023. Acesso em: 18 maio 2024. Disponível em: ⟨https://download.inep.gov.br/acoes_internacionais/pisa/resultados/2022/apresentacao_pisa_2022_brazil.pdf⟩.

4 SCHWARTZMAN, S.; CASTRO, C. de M. Ensino, formação profissional e a questão da mão de obra. Ensaio: Avaliação e Políticas Públicas em Educação, v. 21, n. 80, p. 563–624, jul./set. 2013. Acesso em: 23 maio 2024. Disponível em: ⟨https://www.scielo.br/j/ensaio/a/B8Kb6jfXqvCrfrfpWWr8Wsm/#⟩.

5 VELOSO, F. Educação e mercado de trabalho. 2022. Acesso em: 25 maio 2024. Disponível em: ⟨https://blogdoibre.fgv.br/posts/educacao-e-mercado-de-trabalho⟩.

6 (IBGE). Sistema Nacional de Avaliação da Educação Básica – SAEB. 2023. Acessado em: 08-set-2024. Disponível em: ⟨https://ces.ibge.gov.br/base-de-dados/metadados/inep/sistema-nacional-de-avaliacao-da-educacao-basica-saeb.html⟩.

7 (INEP). Escalas de Proficiência do SAEB. [S.l.], 2019. Disponível em: ⟨https://download.inep.gov.br/educacao_basica/saeb/2019/microdados_saeb_2019/ESCALAS%20DE%20PROFICINCIA/Escalas%20de%20Proficincia%20do%20Saeb.pdf⟩.

8 (INEP). Microdados SAEB 2019. [S.l.], 2020. Disponível em: ⟨https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/saeb⟩.

9 QEDU. IDEB - Brasil. 2023. Acessado em: 08-set-2024. Disponível em: ⟨https://qedu.org.br/brasil/ideb⟩.

10 (INEP). Relatório de Resultados do SAEB 2019: Volume 1 - 5º e 9º Anos do Ensino Fundamental e Séries Finais do Ensino Médio. 2019. Acessado em: 08 set. 2024. Disponível em: ⟨https://download.inep.gov.br/educacao_basica/saeb/2019/resultados/relatorio_de_resultados_do_saeb_2019_volume_1.pdf⟩.

11 RUMSEY, D. Statistics 2 for Dummies. 1. ed. [S.l.]: John Wiley & Sons, Inc., 2009.

12 TUKEY, J. W. Exploratory Data Analysis. Reading, MA: Addison-Wesley, 1977.

13 MORETTIN, P. A.; SINGER, J. M. Estatística e Ciência de Dados. [S.l.]: Editora Blucher, 2021.

*De acordo com a Associação Brasileira de Normas Técnicas (ABNT NBR 6023)

14 JAMES, G. et al. An Introduction to Statistical Learning with Applications in R. New York, NY: Springer, 2013. (Springer Texts in Statistics). ISBN 978-1-4614-7137-0. Disponível em: ⟨https://doi.org/10.1007/978-1-4614-7138-7⟩.

15 THEVAPALAN, A. Decision Trees in Machine Learning Using R: A comprehensive guide to building, visualizing, and interpreting decision tree models with R. 2023. DataCamp Tutorial. Updated June 1, 2023. Disponível em: ⟨https://www.datacamp.com/tutorial/decision-trees-R⟩.

16 LOH, W.-Y. Fifty years of classification and regression trees. International Statistical Review / Revue Internationale de Statistique, [Wiley, International Statistical Institute (ISI)], v. 82, n. 3, p. 329–348, 2014. ISSN 03067734, 17515823. Disponível em: ⟨http://www.jstor.org/stable/43298996⟩.

17 BREIMAN, L. et al. Classification and Regression Trees. [S.l.]: Wadsworth, Belmont, CA, 1984.

18 WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. Climate Research, Inter-Research, v. 30, n. 1, p. 79–82, 2005. ISSN 0936-577X.

19 BOURANY, T. Les 5v du big data. Regards croises sur l'economie, La Découverte, v. 23, n. 2, p. 27–31, 2018.

20 INOUE, G.; BERNARDES, G. Inep aponta piora em todos os níveis da educação básica devido à pandemia. 2022. CNN Brasil, em Brasília, 16 set. 2022. Acesso em: 07 set. 2024. Disponível em: ⟨https://www.cnnbrasil.com.br/nacional/inep-aponta-piora-em-todos-os-niveis-da-educacao-basica-devido-a-pandemia/⟩.

21 Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Indicador de Nível Socioeconômico do SAEB 2019: Nota Técnica. Brasília, Brasil, 2019. Diretoria de Gestão e Planejamento (DGP) e Diretoria de Avaliação da Educação Básica (DAEB). Disponível em: ⟨https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/indicador_nivel_socioeconomico_saeb_2019_nota_tecnica.pdf⟩.

A Anexo - Lista Geral de Variáveis

Tabela 38: Dicionário de variáveis - Parte 1

Parte	Variável	Descrição	Contagem	Tipo
ID	ID_ESCOLA	Máscaras dos Códigos de Escola (são códigos fictícios)	37871	ID
Saída	MEDIA_5EF	Média geral	37871	Num
Informações demográficas	ID_AREA	Área	37871	Char
	ID_LOCALIZACAO	Localização	37871	Char
	ID_REGIAO	Código da Região	37871	Char
	ID_UF	Código da Unidade da Federação	37871	Char
	NIVEL_SOCIO_ECONOMICO	Nível socioeconômico	37871	Char
Informações do diretor	TX_RESP_Q001	Qual é a sua cor ou raça?	37650	Char
	TX_RESP_Q002E3	Por quanto tempo você trabalhou como professor(a) antes de se tornar diretor(a)? Anos	36498	Num
	TX_RESP_Q004E5	Você possui quanto tempo de experiência como diretor(a) de escola? Anos	34566	Num
	TX_RESP_Q006E7	Há quanto tempo você é diretor(a) desta escola? Anos	33177	Num
	TX_RESP_Q008	Considerando todas as suas atividades profissionais remuneradas, quantas horas você trabalha em uma semana normal?	35984	Num
	TX_RESP_Q009	Quantas horas você trabalha em uma semana normal em atividades relacionadas à educação?	35821	Num
	TX_RESP_Q010	Na semana normal de trabalho, quantas horas você trabalha para esta escola?	36033	Num
	TX_RESP_Q011	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Prestação de contas	36115	Num
	TX_RESP_Q012	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Reunião com professores(as)	36492	Num
	TX_RESP_Q013	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Atendimento aos pais ou responsáveis	36119	Num
	TX_RESP_Q014	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Gerenciamento de conflitos	35922	Num
	TX_RESP_Q015	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Atendimento aos(às) alunos(as)	35366	Num
	TX_RESP_Q016	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Atendimento individual aos(às) professores(as)	36159	Num
	TX_RESP_Q017	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Demandas da Secretaria de Educação	36031	Num
	TX_RESP_Q018	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Merenda	36032	Num
	TX_RESP_Q019	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Manutenção	35887	Num
	TX_RESP_Q020	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Segurança	34936	Num
	TX_RESP_Q021	Em uma semana normal de trabalho, quantas horas você costuma gastar, aproximadamente, com as seguintes atividades para esta escola: Outras atividades	33099	Num
	TX_RESP_Q022	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Liderar a equipe escolar.	37363	Char
	TX_RESP_Q023	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Atender as demandas administrativas da rede escolar.	37225	Char
	TX_RESP_Q024	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Atender as demandas administrativas da escola.	37294	Char
	TX_RESP_Q025	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Garantir a manutenção da escola.	37242	Char
	TX_RESP_Q026	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Resolver as demandas dos familiares dos(as) alunos(as).	37267	Char
	TX_RESP_Q027	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Mobilizar a comunidade para auxiliar a escola.	37235	Char
	TX_RESP_Q028	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Coordenar a implantação do Projeto Político-Pedagógico.	37311	Char
	TX_RESP_Q029	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Administrar conflitos.	37216	Char
	TX_RESP_Q030	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Manter os(as) professores(as) motivados(as).	37185	Char
	TX_RESP_Q031	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Avaliar o desempenho dos(as) professores(as).	37230	Char
	TX_RESP_Q032	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Realizar a autoavaliação institucional.	37310	Char
	TX_RESP_Q033	Quanto você sente estar preparado(a) para realizar a seguinte atividade: Melhorar os processos pedagógicos da sua escola.	37433	Char

Tabela 39: Dicionário de variáveis - Parte 2

Parte	Variável	Descrição	Contagem	Tipo
	TX_RESP_Q034	Indique quais são as etapas educacionais atendidas pela sua escola: Educação Infantil e Creche (0 a 3 anos).	31354	Char
	TX_RESP_Q035	Indique quais são as etapas educacionais atendidas pela sua escola: Educação Infantil e Pré-escola (4 e 5 anos).	33699	Char
	TX_RESP_Q036	Indique quais são as etapas educacionais atendidas pela sua escola: Anos Iniciais do Ensino Fundamental.	37748	Char
	TX_RESP_Q037	Indique quais são as etapas educacionais atendidas pela sua escola: Anos Finais do Ensino Fundamental.	33240	Char
	TX_RESP_Q038	Indique quais são as etapas educacionais atendidas pela sua escola: Ensino Médio.	30244	Char
	TX_RESP_Q041	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: Os recursos financeiros foram suficientes.	37787	Char
	TX_RESP_Q042	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: Havia professores(as) para todas as disciplinas.	37747	Char
	TX_RESP_Q043	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: Havia quantidade suficiente de pessoal administrativo.	37745	Char
	TX_RESP_Q044	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: Havia quantidade suficiente de pessoal para apoio pedagógico (coordenador, orientador etc.).	37765	Char
	TX_RESP_Q045	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: Os recursos pedagógicos foram suficientes.	37731	Char
	TX_RESP_Q046	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: Recebi apoio da Secretaria de Educação.	37739	Char
	TX_RESP_Q047	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: Os(As) professores(as) foram assíduos(as).	37702	Char
	TX_RESP_Q048	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: Os(As) professores(as) iniciaram as aulas no horário marcado.	37738	Char
	TX_RESP_Q049	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: As substituições das ausências de professores(as) foram facilmente realizadas.	37749	Char
	TX_RESP_Q050	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: Os(As) alunos(as) foram assíduos(as).	37732	Char
	TX_RESP_Q051	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: Troquei experiências com diretores(as) de outras escolas.	37743	Char
	TX_RESP_Q052	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: A comunidade apoiou a gestão da escola.	37701	Char
	TX_RESP_Q053	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: A comunidade executou trabalhos voluntários na escola.	37702	Char
	TX_RESP_Q054	Indique o quanto concorda ou discorda das afirmativas relativas às condições de funcionamento desta escola neste ano: As famílias contribuíram com o trabalho pedagógico.	37697	Char
Condições de funcionamento da escola	TX_RESP_Q055	Os livros didáticos foram entregues antes do início das aulas	37596	Char
	TX_RESP_Q056	Todos(as) os(as) alunos(as) receberam livros didáticos	37643	Char
	TX_RESP_Q057	O calendário escolar pré-estabelecido foi cumprido sem interrupções?	37690	Char
	TX_RESP_Q068	Sobre os fatos listados abaixo, diga a frequência com que ocorreram neste ano, nesta escola: Profissionais foram vítimas de atentado à vida	37675	Char
	TX_RESP_Q069	Sobre os fatos listados abaixo, diga a frequência com que ocorreram neste ano, nesta escola: Profissionais foram ameaçados(as) por algum aluno	37676	Char
	TX_RESP_Q070	Sobre os fatos listados abaixo, diga a frequência com que ocorreram neste ano, nesta escola: Houve ocorrência de roubo com uso de violência	37660	Char
	TX_RESP_Q071	Sobre os fatos listados abaixo, diga a frequência com que ocorreram neste ano, nesta escola: Houve tráfico de drogas na escola	37656	Char
	TX_RESP_Q072	Sobre os fatos listados abaixo, diga a frequência com que ocorreram neste ano, nesta escola: Alunos(as) frequentaram a escola sob efeito de bebida alcoólica	37591	Char
	TX_RESP_Q073	Sobre os fatos listados abaixo, diga a frequência com que ocorreram neste ano, nesta escola: Alunos(as) frequentaram a escola sob efeito de drogas ilícitas	37617	Char
	TX_RESP_Q074	Sobre os fatos listados abaixo, diga a frequência com que ocorreram neste ano, nesta escola: Alunos(as) frequentaram a escola portando arma (revólver, faca, canivete etc.)	37662	Char
	TX_RESP_Q075	Sobre os fatos listados abaixo, diga a frequência com que ocorreram neste ano, nesta escola: Episódios de violência ocasionaram cancelamento das aulas	37711	Char

Tabela 40: Dicionário de variáveis - Parte 3

Parte	Variável	Descrição	Contagem	Tipo
Recursos e infraestrutura	TX_RESP_Q079	Sobre a área externa da sua escola (pátio, parque e área verde), indique os itens existentes: Bebedouro ao alcance das crianças	37725	Char
	TX_RESP_Q080	Sobre a área externa da sua escola (pátio, parque e área verde), indique os itens existentes: Banheiro infantil	37725	Char
	TX_RESP_Q081	Sobre a área externa da sua escola (pátio, parque e área verde), indique os itens existentes: Área sombreada	37725	Char
	TX_RESP_Q082	Sobre a área externa da sua escola (pátio, parque e área verde), indique os itens existentes: Área coberta	37725	Char
	TX_RESP_Q083	Sobre a área externa da sua escola (pátio, parque e área verde), indique os itens existentes: Vegetação e jardim	37725	Char
	TX_RESP_Q084	Sobre a área externa da sua escola (pátio, parque e área verde), indique os itens existentes: Horta	37725	Char
	TX_RESP_Q085	Quais os tipos de solo ou revestimento do solo da área externa da sua escola? Piso emborrachado	37725	Char
	TX_RESP_Q086	Quais os tipos de solo ou revestimento do solo da área externa da sua escola? Cimento liso	37725	Char
	TX_RESP_Q087	Quais os tipos de solo ou revestimento do solo da área externa da sua escola? Cimento áspero	37725	Char
	TX_RESP_Q088	Quais os tipos de solo ou revestimento do solo da área externa da sua escola? Areia	37725	Char
	TX_RESP_Q089	Quais os tipos de solo ou revestimento do solo da área externa da sua escola? Grama	37725	Char
	TX_RESP_Q090	Quais os tipos de solo ou revestimento do solo da área externa da sua escola? Terra	37725	Char
	TX_RESP_Q091	Quais os tipos de solo ou revestimento do solo da área externa da sua escola? Cerâmica	37725	Char
	TX_RESP_Q092	Quais os tipos de solo ou revestimento do solo da área externa da sua escola? Pedras	37725	Char
	TX_RESP_Q093	Quais equipamentos existem nas áreas externas de recreação da sua escola? Tanque de areia	37725	Char
	TX_RESP_Q094	Quais equipamentos existem nas áreas externas de recreação da sua escola? Gira-gira	37725	Char
	TX_RESP_Q095	Quais equipamentos existem nas áreas externas de recreação da sua escola? Túnel lúdico	37725	Char
	TX_RESP_Q096	Quais equipamentos existem nas áreas externas de recreação da sua escola? Gangorra	37725	Char
	TX_RESP_Q097	Quais equipamentos existem nas áreas externas de recreação da sua escola? Escorregador	37725	Char
	TX_RESP_Q098	Quais equipamentos existem nas áreas externas de recreação da sua escola? Casinha	37725	Char
	TX_RESP_Q099	Quais equipamentos existem nas áreas externas de recreação da sua escola? Balanço	37725	Char
	TX_RESP_Q100	Quais equipamentos existem nas áreas externas de recreação da sua escola? Brinquedo para escalar	37725	Char
	TX_RESP_Q101	Quais equipamentos existem nas áreas externas de recreação da sua escola? Banco	37725	Char
	TX_RESP_Q102	Quais equipamentos existem nas áreas externas de recreação da sua escola? Outros	37725	Char
	TX_RESP_Q103	Avalie os seguintes aspectos da escola: Condições de uso dos equipamentos da área externa de recreação.	33381	Char
	TX_RESP_Q104	Avalie os seguintes aspectos da escola: O acesso à área externa de recreação pelos(as) alunos(as) público-alvo da educação especial.	35234	Char
	TX_RESP_Q105	Avalie os seguintes aspectos da escola: O acesso à entrada principal das pessoas com deficiência física e visual (ex.: rampas e marcadores no chão).	37305	Char
	TX_RESP_Q106	Avalie os seguintes aspectos da escola: Segurança na entrada e saída dos(as) alunos(as) da escola.	37589	Char
	TX_RESP_Q107	Avalie os seguintes aspectos da escola: Muros e grades que impedem que os(as) alunos(as) saiam sozinhos(as).	37634	Char
	TX_RESP_Q108	Avalie os seguintes aspectos da escola: Identificação externa que caracterize o prédio como uma instituição escolar.	37694	Char

Tabela 41: Dicionário de variáveis - Parte 4

Parte	Variável	Descrição	Contagem	Tipo
	TX_RESP_Q118	O Conselho Escolar é um colegiado geralmente constituído por representantes da escola e da comunidade que tem como objetivo acompanhar as atividades escolares. Na sua escola existe Conselho Escolar?	37686	Char
	TX_RESP_Q119	Quantas reuniões do Conselho Escolar ocorreram neste ano?	31007	Num
	TX_RESP_Q120	Considere os atores relacionados a seguir e indique quantos participam do Conselho Escolar: Professores(as)	30936	Num
	TX_RESP_Q121	Considere os atores relacionados a seguir e indique quantos participam do Conselho Escolar: Alunos(as)	19778	Num
	TX_RESP_Q122	Considere os atores relacionados a seguir e indique quantos participam do Conselho Escolar: Pais (ou responsáveis)	30712	Num
	TX_RESP_Q123	Considere os atores relacionados a seguir e indique quantos participam do Conselho Escolar: Funcionários	30109	Num
	TX_RESP_Q124	Considere os atores relacionados a seguir e indique quantos participam do Conselho Escolar: Outros membros	19141	Num
	TX_RESP_Q125	Neste ano, indique a frequência com que os temas/assuntos foram discutidos pelo Conselho Escolar: Questões pedagógicas	31227	Char
	TX_RESP_Q126	Neste ano, indique a frequência com que os temas/assuntos foram discutidos pelo Conselho Escolar: Questões administrativas e institucionais	31199	Char
	TX_RESP_Q127	Neste ano, indique a frequência com que os temas/assuntos foram discutidos pelo Conselho Escolar: Questões financeiras	31141	Char
	TX_RESP_Q128	Neste ano, indique a frequência com que os temas/assuntos foram discutidos pelo Conselho Escolar: Questões de relacionamento com a comunidade	31183	Char
	TX_RESP_Q129	O Conselho Escolar tem função deliberativa?	30976	Char
	TX_RESP_Q130	O Conselho de Classe é um órgão formado por todos os professores que lecionam em cada turma/ano. Neste ano e nesta escola, quantas vezes se reuniu o Conselho de Classe?	30208	Num
	TX_RESP_Q131	O Conselho de Classe é um órgão formado por todos os professores que lecionam em cada turma/ano. Neste ano e nesta escola, quantas vezes se reuniu o Conselho de Classe? Não existe Conselho	37871	Char
	TX_RESP_Q132	A APM - Associação de Pais e Mestres existe para apoiar as ações da escola e integrar a comunidade. Neste ano e nesta escola, quantas vezes se reuniu a APM (ou caixa escolar)?	19378	Num
	TX_RESP_Q133	A APM - Associação de Pais e Mestres existe para apoiar as ações da escola e integrar a comunidade. Neste ano e nesta escola, quantas vezes se reuniu a APM (ou caixa escolar)? Não existe APM.	37871	Char
	TX_RESP_Q134	Há Grêmio Estudantil?	18618	Char
	TX_RESP_Q135	A escola é administrada pela Polícia Militar	37715	Char
	TX_RESP_Q136	Os (As) estudantes são preparados (as) para os testes de avaliação externos.	37700	Char
	TX_RESP_Q137	A escola segue orientação religiosa	37578	Char
	TX_RESP_Q138	Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Programa Dinheiro Direto da Escola.	37680	Char
	TX_RESP_Q139	Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Eventos promovidos nas dependências da escola (Festas, rifas etc.).	37705	Char
	TX_RESP_Q140	Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Empresas que apoiam a escola.	37540	Char
	TX_RESP_Q141	Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Organizações sem fins lucrativos.	37436	Char
Gestão e participação	TX_RESP_Q142	Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Repasses da rede de ensino.	37263	Char
	TX_RESP_Q143	Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Pagamento de taxas pelos familiares dos(as) estudantes.	37500	Char
	TX_RESP_Q144	Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Contribuições voluntárias dos familiares dos(as) estudantes.	37440	Char
	TX_RESP_Q145	Indique se as fontes abaixo fornecem ou não fornecem recursos para o funcionamento desta escola: Contribuições voluntárias dos(as) profissionais da escola.	37482	Char
	TX_RESP_Q146	A escola oferece merenda aos(às) estudantes?	37683	Char
	TX_RESP_Q147	Quantas refeições são oferecidas nesta escola para alunos(as) que permanecem menos de 4 horas na escola	37055	Char
	TX_RESP_Q148	Quantas refeições são oferecidas nesta escola para alunos(as) que permanecem entre 4 e 7 horas na escola	36632	Char
	TX_RESP_Q149	Quantas refeições são oferecidas nesta escola para alunos(as) que permanecem mais de 7 horas na escola	36080	Char
	TX_RESP_Q150	A quantidade de alimentos é suficiente para todos(as)	37460	Char
	TX_RESP_Q151	Os alimentos são de boa qualidade	37423	Char
	TX_RESP_Q152	A cozinha atende todas as necessidades do preparo da merenda	37395	Char
	TX_RESP_Q153	Todos(as) conseguem se alimentar sentados	37427	Char
	TX_RESP_Q154	O acesso ao local de alimentação é livre para pessoas com mobilidade reduzida	37342	Char
	TX_RESP_Q155	Há pias para higienização das mãos próximas ao local de alimentação	37431	Char
	TX_RESP_Q156	A merenda escolar é preparada na própria instituição?	37507	Char
	TX_RESP_Q157	Os cardápios da alimentação escolar são elaborados por nutricionista?	37489	Char

Tabela 42: Dicionário de variáveis - Parte 5

Parte	Variável	Descrição	Contagem	Tipo
Gestão pedagógica	TX_RESP_Q158	A escola possui Projeto Político-Pedagógico?	37689	Char
	TX_RESP_Q159	Seu conteúdo é discutido em reuniões?	35543	Char
	TX_RESP_Q160	Os(As) professores(as) participaram da elaboração?	35536	Char
	TX_RESP_Q161	Os pais participaram da elaboração?	35458	Char
	TX_RESP_Q162	Os(As) estudantes participaram da elaboração?	35430	Char
	TX_RESP_Q163	Estabelece metas de aprendizagem?	35487	Char
	TX_RESP_Q164	Considera os resultados de avaliações externas (Saeb, estaduais, municipais etc.)?	35473	Char
	TX_RESP_Q165	Há metas de alcance de indicadores externos (Ideb, índices estaduais ou municipais)?	35472	Char
	TX_RESP_Q166	Neste ano e nesta escola, todos que solicitaram vagas conseguiram se matricular?	37577	Char
	TX_RESP_Q175	Quais critérios foram considerados para a formação das turmas: Não se aplica	37871	Char
	TX_RESP_Q176	Quais critérios foram considerados para a formação das turmas: Afinidade entre os(as) estudantes	37871	Char
	TX_RESP_Q177	Quais critérios foram considerados para a formação das turmas: Agrupar os(as) estudantes segundo a idade	37871	Char
	TX_RESP_Q178	Quais critérios foram considerados para a formação das turmas: Equilíbrio de meninos e meninas nas turmas	37871	Char
	TX_RESP_Q179	Quais critérios foram considerados para a formação das turmas: Manter as turmas existentes do ano anterior	37871	Char
	TX_RESP_Q180	Quais critérios foram considerados para a formação das turmas: Agrupar os(as) estudantes por critérios disciplinares	37871	Char
	TX_RESP_Q181	Quais critérios foram considerados para a formação das turmas: Agrupar os(as) estudantes com base no seu desempenho	37871	Char
	TX_RESP_Q182	Quais critérios foram considerados para a formação das turmas: Outro	37871	Char
	TX_RESP_Q183	Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Preferência dos(as) professores(as).	37270	Char
	TX_RESP_Q184	Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Tempo de serviço.	37205	Char
	TX_RESP_Q185	Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Cursos de formação continuada realizados.	36984	Char
	TX_RESP_Q186	Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Professores(as) experientes nas turmas com facilidade de aprendizagem.	36827	Char
	TX_RESP_Q187	Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Professores(as) experientes nas turmas com dificuldade de aprendizagem.	37014	Char
	TX_RESP_Q188	Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Manutenção do(a) professor(a) com a mesma turma.	36896	Char
	TX_RESP_Q189	Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Revezamento dos(as) professores(as) entre séries/anos.	36805	Char
	TX_RESP_Q190	Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Sorteio das turmas entre os(as) professores(as).	36657	Char
	TX_RESP_Q191	Neste ano, quais critérios foram utilizados para a atribuição das turmas aos(às) professores(as)? Atribuição pela gestão da escola.	36947	Char
	TX_RESP_Q192	Na sua escola há parcerias com: Ministério Público	37871	Char
	TX_RESP_Q193	Na sua escola há parcerias com: Conselho Tutelar	37871	Char
	TX_RESP_Q194	Na sua escola há parcerias com: Secretaria de Saúde	37871	Char
	TX_RESP_Q195	Na sua escola há parcerias com: Secretaria de Educação	37871	Char
	TX_RESP_Q196	Na sua escola há parcerias com: Secretaria de Assistência Social	37871	Char
	TX_RESP_Q197	Na sua escola há parcerias com: Secretaria de Segurança Pública	37871	Char
	TX_RESP_Q198	Na sua escola há parcerias com: Organizações não governamentais/instituições privadas	37871	Char
	TX_RESP_Q199	Neste ano, para redução do ABANDONO ESCOLAR, avalie o resultado das seguintes ações realizadas nesta escola: Entrar em contato com os familiares do(a) estudante	37767	Char
	TX_RESP_Q200	Neste ano, para redução do ABANDONO ESCOLAR, avalie o resultado das seguintes ações realizadas nesta escola: Ir à residência do(a) estudante	37713	Char
	TX_RESP_Q201	Neste ano, para redução do ABANDONO ESCOLAR, avalie o resultado das seguintes ações realizadas nesta escola: Informar ao Conselho Tutelar	37693	Char
	TX_RESP_Q202	Neste ano, para a redução da REPETÊNCIA ESCOLAR, avalie o resultado das seguintes ações realizadas nesta escola: Oferta de reforço escolar	37685	Char
	TX_RESP_Q203	Neste ano, para a redução da REPETÊNCIA ESCOLAR, avalie o resultado das seguintes ações realizadas nesta escola: Os(As) estudantes são estimulados(as) a apoiar uns(umas) aos(às) outros(as)	37676	Char
	TX_RESP_Q204	Neste ano, para a redução da REPETÊNCIA ESCOLAR, avalie o resultado das seguintes ações realizadas nesta escola: Revisão dos procedimentos de avaliação	37654	Char
	TX_RESP_Q205	Neste ano, para a redução da REPETÊNCIA ESCOLAR, avalie o resultado das seguintes ações realizadas nesta escola: Revisão das práticas pedagógicas	37655	Char

Tabela 43: Dicionário de variáveis - Parte 6

Parte	Variável	Descrição	Contagem	Tipo
Gestão pedagógica	TX_RESP_Q206	Nesta escola, há projetos com as seguintes temáticas: Violência	37871	Char
	TX_RESP_Q207	Nesta escola, há projetos com as seguintes temáticas: Bullying	37871	Char
	TX_RESP_Q208	Nesta escola, há projetos com as seguintes temáticas: Machismo	37871	Char
	TX_RESP_Q209	Nesta escola, há projetos com as seguintes temáticas: Homofobia	37871	Char
	TX_RESP_Q210	Nesta escola, há projetos com as seguintes temáticas: Sexualidade	37871	Char
	TX_RESP_Q211	Nesta escola, há projetos com as seguintes temáticas: Uso de drogas	37871	Char
	TX_RESP_Q212	Nesta escola, há projetos com as seguintes temáticas: Direitos dos idosos	37871	Char
	TX_RESP_Q213	Nesta escola, há projetos com as seguintes temáticas: Educação ambiental	37871	Char
	TX_RESP_Q214	Nesta escola, há projetos com as seguintes temáticas: Ciência e tecnologia	37871	Char
	TX_RESP_Q215	Nesta escola, há projetos com as seguintes temáticas: Diversidade religiosa	37871	Char
	TX_RESP_Q216	Nesta escola, há projetos com as seguintes temáticas: Desigualdades sociais	37871	Char
	TX_RESP_Q217	Nesta escola, há projetos com as seguintes temáticas: Nutrição e alimentação	37871	Char
	TX_RESP_Q218	Nesta escola, há projetos com as seguintes temáticas: Educação para o trânsito	37871	Char
	TX_RESP_Q219	Nesta escola, há projetos com as seguintes temáticas: Relações étnico-raciais/racismo	37871	Char
	TX_RESP_Q220	Nesta escola, há projetos com as seguintes temáticas: Direitos da criança e do adolescente	37871	Char
	TX_RESP_Q221	Nesta escola, há projetos com as seguintes temáticas: Mundo do trabalho (direitos, relações etc)	37871	Char
	TX_RESP_Q222	Nesta escola, há projetos com as seguintes temáticas: Educação financeira e consumo sustentável	37871	Char
	TX_RESP_Q223	Indique se neste ano a escola ofereceu atividades de formação nas seguintes áreas: Conteúdo e compreensão dos conceitos da(s) área(s) de ensino.	37150	Char
	TX_RESP_Q224	Indique se neste ano a escola ofereceu atividades de formação nas seguintes áreas: Avaliação da aprendizagem.	37535	Char
	TX_RESP_Q225	Indique se neste ano a escola ofereceu atividades de formação nas seguintes áreas: Avaliação em larga escala.	36610	Char
	TX_RESP_Q226	Indique se neste ano a escola ofereceu atividades de formação nas seguintes áreas: Metodologias de ensino.	37211	Char
	TX_RESP_Q227	Indique se neste ano a escola ofereceu atividades de formação nas seguintes áreas: Conhecimento do currículo.	37175	Char
	TX_RESP_Q228	Indique se neste ano a escola ofereceu atividades de formação nas seguintes áreas: Gestão da sala de aula.	36783	Char
	TX_RESP_Q229	Indique se neste ano a escola ofereceu atividades de formação nas seguintes áreas: Educação inclusiva.	37089	Char
	TX_RESP_Q230	Indique se neste ano a escola ofereceu atividades de formação nas seguintes áreas: Novas tecnologias.	36298	Char
	TX_RESP_Q231	Indique se neste ano a escola ofereceu atividades de formação nas seguintes áreas: Gestão e administração escolar.	36604	Char
Educação inclusiva	TX_RESP_Q232	Os(As) profissionais para inclusão do público-alvo da educação especial são em número suficiente?	37612	Char
	TX_RESP_Q241	Nos últimos doze meses, sua escola recebeu treinamento para lidar com o público-alvo da educação especial?	37691	Char

B Anexo - Código Geral em R

```r
#Codigo Geral Adaptavel

library(openxlsx)     # leitura de excel
library(rsample)      # criacao de amostras
library(rpart)        # arvores de regressao
library(rpart.plot)   # plotagem arvores
library(caret)        # atributos da arvore
library(Metrics)      # calculo de erro

# 1. LEITURA DE BASE E SEPARACAO

base <- read.xlsx("C:/Users/Usuario/Downloads/versao resumida base tf 2.xlsx", sheet = "diretor r", startRow = 2, colNames = TRUE)

#View(base)
set.seed(123)
indices <- sample(1:nrow(base), floor(0.8*nrow(base)), replace = FALSE)
conjunto_treino <- base[indices, ]
conjunto_teste <- base[-indices, ]

#2. MODELAGEM

modelo_arvore <- rpart( MEDIA_5F ~

#GERAL 1234567 (geral)
#ID_AREA + ID_LOCALIZACAO + ID_REGIAO + ID_UF + NIVEL_SOCIO_ECONOMICO + TX_RESP_Q001 + TX_RESP_Q002E3 + TX_RESP_Q004E5 + TX_RESP_Q006E7 + TX_RESP_Q008 + TX_RESP_Q009 + TX_RESP_Q010 + TX_RESP_Q011 + TX_RESP_Q012 + TX_RESP_Q013 + TX_RESP_Q014 + TX_RESP_Q015 + TX_RESP_Q016 + TX_RESP_Q017 + TX_RESP_Q018 + TX_RESP_Q019 + TX_RESP_Q020 + TX_RESP_Q021 + TX_RESP_Q022 + TX_
```

```
RESP_Q023 + TX_RESP_Q024 + TX_RESP_Q025 + TX_RESP_Q026 + TX_
RESP_Q027 + TX_RESP_Q028 + TX_RESP_Q029 + TX_RESP_Q030 + TX_
RESP_Q031 + TX_RESP_Q032 + TX_RESP_Q033 + TX_RESP_Q034 + TX_
RESP_Q035 + TX_RESP_Q036 + TX_RESP_Q037 + TX_RESP_Q038 + TX_
RESP_Q041 + TX_RESP_Q042 + TX_RESP_Q043 + TX_RESP_Q044 + TX_
RESP_Q045 + TX_RESP_Q046 + TX_RESP_Q047 + TX_RESP_Q048 + TX_
RESP_Q049 + TX_RESP_Q050 + TX_RESP_Q051 + TX_RESP_Q052 + TX_
RESP_Q053 + TX_RESP_Q054 + TX_RESP_Q055 + TX_RESP_Q056 + TX_
RESP_Q057 + TX_RESP_Q068 + TX_RESP_Q069 + TX_RESP_Q070 + TX_
RESP_Q071 + TX_RESP_Q072 + TX_RESP_Q073 + TX_RESP_Q074 + TX_
RESP_Q075 + TX_RESP_Q079 + TX_RESP_Q080 + TX_RESP_Q081 + TX_
RESP_Q082 + TX_RESP_Q083 + TX_RESP_Q084 + TX_RESP_Q085 + TX_
RESP_Q086 + TX_RESP_Q087 + TX_RESP_Q088 + TX_RESP_Q089 + TX_
RESP_Q090 + TX_RESP_Q091 + TX_RESP_Q092 + TX_RESP_Q093 + TX_
RESP_Q094 + TX_RESP_Q095 + TX_RESP_Q096 + TX_RESP_Q097 + TX_
RESP_Q098 + TX_RESP_Q099 + TX_RESP_Q100 + TX_RESP_Q101 + TX_
RESP_Q102 + TX_RESP_Q103 + TX_RESP_Q104 + TX_RESP_Q105 + TX_
RESP_Q106 + TX_RESP_Q107 + TX_RESP_Q108 + TX_RESP_Q118 + TX_
RESP_Q119 + TX_RESP_Q120 + TX_RESP_Q121 + TX_RESP_Q122 + TX_
RESP_Q123 + TX_RESP_Q124 + TX_RESP_Q125 + TX_RESP_Q126 + TX_
RESP_Q127 + TX_RESP_Q128 + TX_RESP_Q129 + TX_RESP_Q130 + TX_
RESP_Q131 + TX_RESP_Q132 + TX_RESP_Q133 + TX_RESP_Q134 + TX_
RESP_Q135 + TX_RESP_Q136 + TX_RESP_Q137 + TX_RESP_Q138 + TX_
RESP_Q139 + TX_RESP_Q140 + TX_RESP_Q141 + TX_RESP_Q142 + TX_
RESP_Q143 + TX_RESP_Q144 + TX_RESP_Q145 + TX_RESP_Q146 + TX_
RESP_Q147 + TX_RESP_Q148 + TX_RESP_Q149 + TX_RESP_Q150 + TX_
RESP_Q151 + TX_RESP_Q152 + TX_RESP_Q153 + TX_RESP_Q154 + TX_
RESP_Q155 + TX_RESP_Q156 + TX_RESP_Q157 + TX_RESP_Q158 + TX_
RESP_Q159 + TX_RESP_Q160 + TX_RESP_Q161 + TX_RESP_Q162 + TX_
RESP_Q163 + TX_RESP_Q164 + TX_RESP_Q165 + TX_RESP_Q166 + TX_
RESP_Q175 + TX_RESP_Q176 + TX_RESP_Q177 + TX_RESP_Q178 + TX_
RESP_Q179 + TX_RESP_Q180 + TX_RESP_Q181 + TX_RESP_Q182 + TX_
RESP_Q183 + TX_RESP_Q184 + TX_RESP_Q185 + TX_RESP_Q186 + TX_
RESP_Q187 + TX_RESP_Q188 + TX_RESP_Q189 + TX_RESP_Q190 + TX_
```

```
   RESP_Q191 + TX_RESP_Q192 + TX_RESP_Q193 + TX_RESP_Q194 + TX_
   RESP_Q195 + TX_RESP_Q196 + TX_RESP_Q197 + TX_RESP_Q198 + TX_
   RESP_Q199 + TX_RESP_Q200 + TX_RESP_Q201 + TX_RESP_Q202 + TX_
   RESP_Q203 + TX_RESP_Q204 + TX_RESP_Q205 + TX_RESP_Q206 + TX_
   RESP_Q207 + TX_RESP_Q208 + TX_RESP_Q209 + TX_RESP_Q210 + TX_
   RESP_Q211 + TX_RESP_Q212 + TX_RESP_Q213 + TX_RESP_Q214 + TX_
   RESP_Q215 + TX_RESP_Q216 + TX_RESP_Q217 + TX_RESP_Q218 + TX_
   RESP_Q219 + TX_RESP_Q220 + TX_RESP_Q221 + TX_RESP_Q222 + TX_
   RESP_Q223 + TX_RESP_Q224 + TX_RESP_Q225 + TX_RESP_Q226 + TX_
   RESP_Q227 + TX_RESP_Q228 + TX_RESP_Q229 + TX_RESP_Q230 + TX_
   RESP_Q231 + TX_RESP_Q232 + TX_RESP_Q241

#GERAL 234567 (geral diretor)
#TX_RESP_Q001 + TX_RESP_Q002E3 + TX_RESP_Q004E5 + TX_RESP_Q006E7
   + TX_RESP_Q008 + TX_RESP_Q009 + TX_RESP_Q010 + TX_RESP_Q011 +
   TX_RESP_Q012 + TX_RESP_Q013 + TX_RESP_Q014 + TX_RESP_Q015 + TX_
   RESP_Q016 + TX_RESP_Q017 + TX_RESP_Q018 + TX_RESP_Q019 + TX_
   RESP_Q020 + TX_RESP_Q021 + TX_RESP_Q022 + TX_RESP_Q023 + TX_
   RESP_Q024 + TX_RESP_Q025 + TX_RESP_Q026 + TX_RESP_Q027 + TX_
   RESP_Q028 + TX_RESP_Q029 + TX_RESP_Q030 + TX_RESP_Q031 + TX_
   RESP_Q032 + TX_RESP_Q033 + TX_RESP_Q034 + TX_RESP_Q035 + TX_
   RESP_Q036 + TX_RESP_Q037 + TX_RESP_Q038 + TX_RESP_Q041 + TX_
   RESP_Q042 + TX_RESP_Q043 + TX_RESP_Q044 + TX_RESP_Q045 + TX_
   RESP_Q046 + TX_RESP_Q047 + TX_RESP_Q048 + TX_RESP_Q049 + TX_
   RESP_Q050 + TX_RESP_Q051 + TX_RESP_Q052 + TX_RESP_Q053 + TX_
   RESP_Q054 + TX_RESP_Q055 + TX_RESP_Q056 + TX_RESP_Q057 + TX_
   RESP_Q068 + TX_RESP_Q069 + TX_RESP_Q070 + TX_RESP_Q071 + TX_
   RESP_Q072 + TX_RESP_Q073 + TX_RESP_Q074 + TX_RESP_Q075 + TX_
   RESP_Q079 + TX_RESP_Q080 + TX_RESP_Q081 + TX_RESP_Q082 + TX_
   RESP_Q083 + TX_RESP_Q084 + TX_RESP_Q085 + TX_RESP_Q086 + TX_
   RESP_Q087 + TX_RESP_Q088 + TX_RESP_Q089 + TX_RESP_Q090 + TX_
   RESP_Q091 + TX_RESP_Q092 + TX_RESP_Q093 + TX_RESP_Q094 + TX_
   RESP_Q095 + TX_RESP_Q096 + TX_RESP_Q097 + TX_RESP_Q098 + TX_
   RESP_Q099 + TX_RESP_Q100 + TX_RESP_Q101 + TX_RESP_Q102 + TX_
```

```
   RESP_Q103 + TX_RESP_Q104 + TX_RESP_Q105 + TX_RESP_Q106 + TX_
   RESP_Q107 + TX_RESP_Q108 + TX_RESP_Q118 + TX_RESP_Q119 + TX_
   RESP_Q120 + TX_RESP_Q121 + TX_RESP_Q122 + TX_RESP_Q123 + TX_
   RESP_Q124 + TX_RESP_Q125 + TX_RESP_Q126 + TX_RESP_Q127 + TX_
   RESP_Q128 + TX_RESP_Q129 + TX_RESP_Q130 + TX_RESP_Q131 + TX_
   RESP_Q132 + TX_RESP_Q133 + TX_RESP_Q134 + TX_RESP_Q135 + TX_
   RESP_Q136 + TX_RESP_Q137 + TX_RESP_Q138 + TX_RESP_Q139 + TX_
   RESP_Q140 + TX_RESP_Q141 + TX_RESP_Q142 + TX_RESP_Q143 + TX_
   RESP_Q144 + TX_RESP_Q145 + TX_RESP_Q146 + TX_RESP_Q147 + TX_
   RESP_Q148 + TX_RESP_Q149 + TX_RESP_Q150 + TX_RESP_Q151 + TX_
   RESP_Q152 + TX_RESP_Q153 + TX_RESP_Q154 + TX_RESP_Q155 + TX_
   RESP_Q156 + TX_RESP_Q157 + TX_RESP_Q158 + TX_RESP_Q159 + TX_
   RESP_Q160 + TX_RESP_Q161 + TX_RESP_Q162 + TX_RESP_Q163 + TX_
   RESP_Q164 + TX_RESP_Q165 + TX_RESP_Q166 + TX_RESP_Q175 + TX_
   RESP_Q176 + TX_RESP_Q177 + TX_RESP_Q178 + TX_RESP_Q179 + TX_
   RESP_Q180 + TX_RESP_Q181 + TX_RESP_Q182 + TX_RESP_Q183 + TX_
   RESP_Q184 + TX_RESP_Q185 + TX_RESP_Q186 + TX_RESP_Q187 + TX_
   RESP_Q188 + TX_RESP_Q189 + TX_RESP_Q190 + TX_RESP_Q191 + TX_
   RESP_Q192 + TX_RESP_Q193 + TX_RESP_Q194 + TX_RESP_Q195 + TX_
   RESP_Q196 + TX_RESP_Q197 + TX_RESP_Q198 + TX_RESP_Q199 + TX_
   RESP_Q200 + TX_RESP_Q201 + TX_RESP_Q202 + TX_RESP_Q203 + TX_
   RESP_Q204 + TX_RESP_Q205 + TX_RESP_Q206 + TX_RESP_Q207 + TX_
   RESP_Q208 + TX_RESP_Q209 + TX_RESP_Q210 + TX_RESP_Q211 + TX_
   RESP_Q212 + TX_RESP_Q213 + TX_RESP_Q214 + TX_RESP_Q215 + TX_
   RESP_Q216 + TX_RESP_Q217 + TX_RESP_Q218 + TX_RESP_Q219 + TX_
   RESP_Q220 + TX_RESP_Q221 + TX_RESP_Q222 + TX_RESP_Q223 + TX_
   RESP_Q224 + TX_RESP_Q225 + TX_RESP_Q226 + TX_RESP_Q227 + TX_
   RESP_Q228 + TX_RESP_Q229 + TX_RESP_Q230 + TX_RESP_Q231 + TX_
   RESP_Q232 + TX_RESP_Q241

29
30 #DIRETOR 2 (tematica 1)
31 #TX_RESP_Q001 + TX_RESP_Q002E3 + TX_RESP_Q004E5 + TX_RESP_Q006E7
   + TX_RESP_Q008 + TX_RESP_Q009 + TX_RESP_Q010 + TX_RESP_Q011 +
   TX_RESP_Q012 + TX_RESP_Q013 + TX_RESP_Q014 + TX_RESP_Q015 + TX_
```

```
    RESP_Q016 + TX_RESP_Q017 + TX_RESP_Q018 + TX_RESP_Q019 + TX_
    RESP_Q020 + TX_RESP_Q021 + TX_RESP_Q022 + TX_RESP_Q023 + TX_
    RESP_Q024 + TX_RESP_Q025 + TX_RESP_Q026 + TX_RESP_Q027 + TX_
    RESP_Q028 + TX_RESP_Q029 + TX_RESP_Q030 + TX_RESP_Q031 + TX_
    RESP_Q032 + TX_RESP_Q033
32
33  #FUNCIONAMENTO 3 e INFRAESTRUTURA 4 (tematica 2)
34  #TX_RESP_Q034 + TX_RESP_Q035 + TX_RESP_Q036 + TX_RESP_Q037 + TX_
    RESP_Q038 + TX_RESP_Q041 + TX_RESP_Q042 + TX_RESP_Q043 + TX_
    RESP_Q044 + TX_RESP_Q045 + TX_RESP_Q046 + TX_RESP_Q047 + TX_
    RESP_Q048 + TX_RESP_Q049 + TX_RESP_Q050 + TX_RESP_Q051 + TX_
    RESP_Q052 + TX_RESP_Q053 + TX_RESP_Q054 + TX_RESP_Q055 + TX_
    RESP_Q056 + TX_RESP_Q057 + TX_RESP_Q068 + TX_RESP_Q069 + TX_
    RESP_Q070 + TX_RESP_Q071 + TX_RESP_Q072 + TX_RESP_Q073 + TX_
    RESP_Q074 + TX_RESP_Q075 + TX_RESP_Q079 + TX_RESP_Q080 + TX_
    RESP_Q081 + TX_RESP_Q082 + TX_RESP_Q083 + TX_RESP_Q084 + TX_
    RESP_Q085 + TX_RESP_Q086 + TX_RESP_Q087 + TX_RESP_Q088 + TX_
    RESP_Q089 + TX_RESP_Q090 + TX_RESP_Q091 + TX_RESP_Q092 + TX_
    RESP_Q093 + TX_RESP_Q094 + TX_RESP_Q095 + TX_RESP_Q096 + TX_
    RESP_Q097 + TX_RESP_Q098 + TX_RESP_Q099 + TX_RESP_Q100 + TX_
    RESP_Q101 + TX_RESP_Q102 + TX_RESP_Q103 + TX_RESP_Q104 + TX_
    RESP_Q105 + TX_RESP_Q106 + TX_RESP_Q107 + TX_RESP_Q108
35
36  #GESTAO 5 PEDAGOGICA 6 E INCLUSIVA 7 (tematica 3)
37  #TX_RESP_Q118 + TX_RESP_Q119 + TX_RESP_Q120 + TX_RESP_Q121 + TX_
    RESP_Q122 + TX_RESP_Q123 + TX_RESP_Q124 + TX_RESP_Q125 + TX_
    RESP_Q126 + TX_RESP_Q127 + TX_RESP_Q128 + TX_RESP_Q129 + TX_
    RESP_Q130 + TX_RESP_Q131 + TX_RESP_Q132 + TX_RESP_Q133 + TX_
    RESP_Q134 + TX_RESP_Q135 + TX_RESP_Q136 + TX_RESP_Q137 + TX_
    RESP_Q138 + TX_RESP_Q139 + TX_RESP_Q140 + TX_RESP_Q141 + TX_
    RESP_Q142 + TX_RESP_Q143 + TX_RESP_Q144 + TX_RESP_Q145 + TX_
    RESP_Q146 + TX_RESP_Q147 + TX_RESP_Q148 + TX_RESP_Q149 + TX_
    RESP_Q150 + TX_RESP_Q151 + TX_RESP_Q152 + TX_RESP_Q153 + TX_
    RESP_Q154 + TX_RESP_Q155 + TX_RESP_Q156 + TX_RESP_Q157 + TX_
```

```
   RESP_Q158 + TX_RESP_Q159 + TX_RESP_Q160 + TX_RESP_Q161 + TX_
   RESP_Q162 + TX_RESP_Q163 + TX_RESP_Q164 + TX_RESP_Q165 + TX_
   RESP_Q166 + TX_RESP_Q175 + TX_RESP_Q176 + TX_RESP_Q177 + TX_
   RESP_Q178 + TX_RESP_Q179 + TX_RESP_Q180 + TX_RESP_Q181 + TX_
   RESP_Q182 + TX_RESP_Q183 + TX_RESP_Q184 + TX_RESP_Q185 + TX_
   RESP_Q186 + TX_RESP_Q187 + TX_RESP_Q188 + TX_RESP_Q189 + TX_
   RESP_Q190 + TX_RESP_Q191 + TX_RESP_Q192 + TX_RESP_Q193 + TX_
   RESP_Q194 + TX_RESP_Q195 + TX_RESP_Q196 + TX_RESP_Q197 + TX_
   RESP_Q198 + TX_RESP_Q199 + TX_RESP_Q200 + TX_RESP_Q201 + TX_
   RESP_Q202 + TX_RESP_Q203 + TX_RESP_Q204 + TX_RESP_Q205 + TX_
   RESP_Q206 + TX_RESP_Q207 + TX_RESP_Q208 + TX_RESP_Q209 + TX_
   RESP_Q210 + TX_RESP_Q211 + TX_RESP_Q212 + TX_RESP_Q213 + TX_
   RESP_Q214 + TX_RESP_Q215 + TX_RESP_Q216 + TX_RESP_Q217 + TX_
   RESP_Q218 + TX_RESP_Q219 + TX_RESP_Q220 + TX_RESP_Q221 + TX_
   RESP_Q222 + TX_RESP_Q223 + TX_RESP_Q224 + TX_RESP_Q225 + TX_
   RESP_Q226 + TX_RESP_Q227 + TX_RESP_Q228 + TX_RESP_Q229 + TX_
   RESP_Q230 + TX_RESP_Q231 + TX_RESP_Q232 + TX_RESP_Q241
38
39 , data = conjunto_treino, method = "anova",
40 control = list(cp=0.001,minsplit = 1000,minbucket=1000, maxdepth
   = 6, xval = 10))
41
42 rpart.plot(fallen.leaves = FALSE, modelo_arvore, compress= FALSE,
    cex = 0.4,
43           type = 5, faclen= -20, extra =101)
44
45 print(modelo_arvore)
46 plotcp(modelo_arvore)
47
48 # 3. TESTE DE PRECISAO DA PREVISAO
49
50 previsao <- predict(modelo_arvore, newdata = conjunto_teste)
51
52 mae <- mae(conjunto_teste$MEDIA_5F, previsao)
```

```
53 mse <- mse(conjunto_teste$MEDIA_5F, previsao)
54 rmse <- rmse(conjunto_teste$MEDIA_5F, previsao)
55 r2 <- 1 - sum((conjunto_teste$MEDIA_5F - previsao)^2) / sum((
       conjunto_teste$MEDIA_5F - mean(conjunto_teste$MEDIA_5F))^2)
56
57 cat("MAE: ", mae, "\nMSE: ", mse, "\nRMSE: ", rmse, "\nR2: ", r2,
       "\n")
```