

Um Estudo Comparativo da Análise dos Erros nos Descritores do Spaece Entre Fortaleza e o Interior do Ceará com Uso do Algoritmo FP-Growth

Gustavo Barreto de Souza, Universidade Federal do Ceará, gus.barretogg@gmail.com
<https://orcid.org/0009-0002-6583-2193>

Marcos Vinicius de Andrade Lima, Universidade Federal do Ceará, marcos.vinicius@ufc.br
<https://orcid.org/0000-0002-5797-0222>

Resumo: Este estudo investigou os padrões de erro em Matemática entre estudantes do 9º ano do ensino fundamental na avaliação em larga escala SPAECE para o ano de 2019, com o objetivo de comparar as dificuldades dos estudantes da capital Fortaleza e os municípios do interior do Ceará. Foram utilizadas técnicas da Ciência de Dados, com a aplicação do algoritmo FP-Growth, visando identificar padrões de dificuldades dos alunos. A análise comparativa entre os desempenhos dos estudantes revelou dificuldades comuns de aprendizado em geometria e em operações com números racionais e constatou que os alunos da capital tendem a cometer mais erros em conteúdos mais complexos, enquanto os do interior demonstram limitações em tópicos mais elementares. Esses achados evidenciam a necessidade de estratégias pedagógicas diferenciadas para o enfrentamento das defasagens de aprendizagem em Matemática.

Palavras-chave: avaliação em larga escala, mineração de dados, ciência de dados, fp-growth, spaece.

A Comparative Study of Error Analysis in SPAECE Descriptors Between Fortaleza and the Interior of Ceará Using the FP-Growth Algorithm

Abstract: This study investigated error patterns in Mathematics among 9th-grade elementary school students in the 2019 SPAECE large-scale assessment, aiming to compare the difficulties faced by students from Fortaleza and those from municipalities in the interior of Ceará. Data Science techniques were employed, specifically the FP-Growth algorithm, to identify patterns of student difficulties. The comparative analysis of student performance revealed common learning challenges in geometry and operations with rational numbers. Furthermore, it was found that students from the capital tend to make more mistakes in more complex content, while those from the interior show limitations in more elementary topics. These findings highlight the need for differentiated pedagogical strategies to address learning gaps in Mathematics.

Keywords: large-Scale assessment, data mining, data science, fp-growth, spaece.

1. Introdução

De acordo com Bauer; Gatti e Tavares (2013) a constante preocupação com a qualidade do ensino público no Brasil tem ganho visibilidade nas políticas educacionais, impulsionada pela consolidação de sistemas de avaliação em larga escala, que visam contribuir para a elevação dos padrões educacionais e garantir o direito de todos a uma educação de qualidade. Essa valorização do ensino reflete-se também em outras áreas, já que, conforme argumenta Grossman (2005), a educação exerce impacto direto em diversas esferas da vida do indivíduo, como a saúde, a economia e até mesmo sua própria subsistência.

As avaliações em larga escala, como o Sistema Permanente de Avaliação da Educação Básica do Ceará (SPAECE), surgem como ferramentas importantes para avaliar o nível de aprendizado dos alunos e identificar áreas que necessitam de melhorias e aquelas que já estão bem consolidadas no aprendizado dos estudantes. O SPAECE,

implementado desde 1992, avalia as competências e habilidades dos estudantes da educação básica em Língua Portuguesa e Matemática, fornecendo dados que orientam políticas educacionais e práticas pedagógicas (Ceará, 2020).

O SPAECE, no decorrer dos anos, vem avaliando a evolução do aprendizado no estado do Ceará. Contribuindo para o desenvolvimento da educação, no qual, no ano de 2023, avaliou 100% dos estudantes do ensino fundamental, matriculados no 2º, 5º e 9º anos da rede pública municipal, reunindo mais de 285 mil alunos distribuídos em 3.695 escolas. Pelos resultados obtidos, é perceptível o crescimento da alfabetização, que alcançou 97% das crianças do 2º ano. Os dados informam que grande parte dos municípios do Ceará atingiu o padrão almejado nas disciplinas de Português e Matemática (Ceará, 2024).

A tese de Lima (2023) avaliou a formação continuada de professores e, por meio de técnicas da Ciência de Dados, identificou dificuldades dos alunos a partir da análise dos resultados do SPAECE do 9º ano do ensino fundamental da rede pública municipal da cidade de Fortaleza, referentes ao ano de 2019. O autor aplicou o algoritmo FP-Growth para verificar padrões de associação entre os erros dos estudantes, a fim de obter percepções de causa e efeito entre os principais descritores errados, possibilitando um plano de melhorias mais apurado para a formação continuada dos professores.

Diante desse panorama, surge a pergunta central desta pesquisa: os resultados do SPAECE para Matemática do 9º ano nos municípios do interior do Ceará demonstram as mesmas dificuldades observadas pelos estudantes de Fortaleza? Essa indagação é fundamental para entender se as dificuldades enfrentadas pelos alunos são uma característica comum em diferentes contextos ou se há variações significativas que possam influenciar as estratégias e políticas educacionais.

Com base na contextualização apresentada e na questão de pesquisa, esse trabalho procura realizar uma análise comparativa entre os resultados obtidos pelos estudantes de Fortaleza e do interior do Ceará na avaliação em larga escala SPAECE para a prova de Matemática do 9º ano do ensino fundamental, para possibilitar uma compreensão mais aprofundada do cenário educacional do estado do Ceará, contribuindo para decisões baseadas em dados, além de poder contribuir para o desenvolvimento de políticas educacionais alinhadas ao contexto local.

O presente trabalho está estruturado em cinco seções principais. A primeira traz a Introdução, que aborda a contextualização do trabalho, apresenta a questão de pesquisa e estabelece seu objetivo. Na segunda seção, temos a fundamentação teórica, nos quais são abordados os temas de avaliação em larga escala e a ciência de dados, em especial a técnica utilizada para descoberta de padrões de associação entre itens. A terceira seção traz os procedimentos metodológicos, com a definição do paradigma e da abordagem de pesquisa utilizados, a especificação das fases desenvolvidas, além dos aspectos éticos da pesquisa. Já na quarta seção, os resultados são apresentados e analisados. A quinta e última seção tece as considerações finais, explicitando nossas descobertas, além de propostas para trabalhos futuros.

2. Referencial Teórico

2.1. Avaliação em larga escala

As primeiras medições do desempenho escolar por meio de testes acadêmicos tiveram seu início no Brasil a partir de meados da década de 60 (Gatti, 2013), contudo, somente no ano de 1988 é que se dá início às experiências de avaliação em larga escala na Educação Básica. Desde então, passaram por diversas reformulações ao longo dos anos (Werle, 2011).

Os objetivos da avaliação em larga escala do sistema escolar são os de informar o que populações e subpopulações de alunos em diferentes séries sabem e são capazes de fazer, em um determinado momento, e acompanhar sua evolução ao longo dos anos. É importante destacar que não é objetivo da avaliação em larga escala fornecer informações sobre alunos ou escolas individuais (Klein e Fontanive, 1995).

A primeira avaliação em larga escala implementada no Brasil foi o Sistema de Avaliação da Educação Básica (SAEB), principal sistema de avaliação da qualidade da educação básica nacional. Essa avaliação ocorre a cada dois anos, na qual, em sua primeira edição, realizou a avaliação com uma amostra dos alunos do 6º e 9º ano do ensino fundamental e no 3º ano do ensino médio, em escolas públicas e privadas das áreas urbana e rural, visando avaliar o desempenho dos alunos nas disciplinas de Língua Portuguesa e Matemática (Bonamino e Sousa, 2012). Desde então, as avaliações em larga escala se tornaram uma prática comum no Brasil, realizadas regularmente pelo Ministério da Educação (MEC) e pelas secretarias de educação dos estados e municípios (Pontes, 2023).

O estado do Ceará, por meio da Secretaria da Educação (SEDUC), desde 1992 vem implementando o SPAECE, que é uma avaliação externa de larga escala destinada a avaliar as competências e habilidades dos alunos do Ensino Fundamental e do Ensino Médio em Língua Portuguesa e Matemática. A avaliação abrange as escolas estaduais e municipais de todos os 184 municípios cearenses e tem como orientação as Matrizes de Referência alinhadas com as do SAEB. As informações coletadas a cada avaliação identificam o nível de proficiência e a evolução do desempenho dos alunos do estado do Ceará (Ceará, 2020).

O SPAECE avalia anualmente, de forma censitária, os alunos regularmente matriculados nas séries iniciais do ensino fundamental (5º ano), séries finais do ensino fundamental (9º ano), ensino médio (3ª série) e educação de jovens e adultos (EJA), níveis fundamental e médio. Além disso, a SEDUC ampliou a avaliação do SPAECE, que a partir do ano de 2007 passou a avaliar o grau de alfabetização dos alunos regularmente matriculados no 2º ano do ensino fundamental, denominado de SPAECE-ALFA (FILHO; VIDAL; JÚNIOR, 2020).

O conjunto de dados coletados pelo SPAECE permite analisar os resultados da aprendizagem dos alunos, seus pontos fracos e fortes, e sobre as características dos professores e gestores das escolas estaduais. Ademais, uma vez que a avaliação é aplicada em diferentes fases da vida de cada estudante, é viável monitorar o progresso e evolução de cada estudante de forma mais individual, ao longo do tempo. Esse sistema tem por objetivo fornecer subsídios para formulação, reformulação e monitoramento das políticas educacionais, além de possibilitar aos professores, dirigentes escolares e gestores um quadro da situação da Educação Básica da Rede Pública de ensino de suas escolas e do estado como um todo (Ceará, 2020). Avaliações em larga escala, como o SPAECE, geram uma grande quantidade de dados. Para análise desses dados, pode-se recorrer à Ciência de Dados, detalha a seguir.

2.2. Dados e Ciência de Dados

De acordo com Silva; Peres e Boscarioli (2016) os dados são fatos, gerados de alguma medição ou de algum valor documentado. Quando atribuímos um sentido ou significado aos dados, conseguimos a informação. Quando essa informação é aprendida pelo indivíduo e ele se torna consciente e capaz de tomar decisões relacionadas a ela, temos o conhecimento.

Com o crescimento da Web e a popularização dos sistemas em nuvem, vivemos

uma era marcada por uma crescente geração e armazenamento de dados digitais. Segundo Bérghamo e Oliveira (2019), o volume de dados armazenados no mundo digital cresce a cada ano, impulsionado por atividades cotidianas como navegação na internet, interações em redes sociais, preenchimento de formulários e realização de testes e avaliações.

A Ciência de Dados, de acordo com Saldanha; Barcellos e Pedroso (2021), é um campo de estudo que se destaca pela capacidade de auxiliar a descoberta de informação útil a partir de grandes ou complexas bases de dados, bem como a tomada de decisão orientada por dados, como: estratégias para extração, transformação e carga dos dados, modelagem, construção e avaliação de algoritmos descritivos e preditivos, visualização de grandes quantidades de dados e *deploy* dos modelos em ambientes de produção para a tomada de decisão, entre outros.

A Ciência de Dados, para cumprir com seus propósitos, utiliza-se de métodos de mineração de dados compostos por várias técnicas criadas para facilitar a descoberta de algo novo que possa contribuir no conhecimento a partir de milhares de informações (Lorenzetti e Telöcken, 2016). Entre as técnicas disponíveis e adotadas por nós nesta pesquisa, temos as regras de associação entre itens.

2.2.1. Regras de associação entre itens

Uma regra de associação é uma expressão que relaciona dois conjuntos de itens A e B. Isso significa que, se uma transação em uma base de dados contém A, ela tende também a conter B. O conjunto de itens à esquerda da seta $A \rightarrow B$ é chamado de antecedente, e o conjunto de itens à direita da seta é chamado de consequente (Mariano, 2011).

Para cada regra de associação extraída de uma base de dados, são atribuídos ainda dois fatores importantes: suporte e confiança. O suporte refere-se à probabilidade com que os itens de A e B aparecem juntos nas transações da base de dados. Já a confiança refere-se à probabilidade de encontrar os itens de B em uma transação, dado que essa transação já contém os itens de A, ou seja, é uma medida de quão frequentemente os itens do consequente aparecem nas transações que contêm o antecedente (Mariano, 2011).

Uma regra é considerada frequente se o suporte dela é maior ou igual a um valor mínimo especificado pelo usuário no algoritmo, além disso, uma regra é considerada forte se a confiança dela é maior ou igual a um valor mínimo (confiança mínima) também especificado pelo usuário (Mariano, 2011).

2.2.2. FP-Growth

Entre os algoritmos que implementam a descoberta de padrões entre os itens, utilizamos nesta pesquisa o FP-Growth, que adota uma abordagem que codifica o conjunto de dados em uma estrutura de dados chamada FP-tree ou árvore de padrões frequentes e extrai os conjuntos de itens frequentes e as regras de associação diretamente desta estrutura. O algoritmo tem o critério de ordenação por popularidade decrescente, ou seja, o primeiro item é o mais popular na base de dados e o último item na ordenação é o menos popular (Prado *et al.*, 2002).

O algoritmo FP-Growth pode ser dividido em duas fases. Na primeira fase, a frequência de cada item na base de dados é determinada a partir da representação altamente concentrada da base de dados, o FP-Tree, que é feita através da estratégia de busca em profundidade. Na segunda fase, o FP-Growth utiliza a FP-Tree para determinar os valores de suporte e confiança para todas as regras geradas do conjunto de itens frequentes (Mariano, 2011). O algoritmo também tem como saída o valor lift para cada padrão de associação identificado. Um *lift* igual a um significa que não há relação no

conjunto; um *lift* maior que um significa que existe uma relação positiva no conjunto; e um *lift* menor que um, há uma relação negativa no conjunto. Em nossa pesquisa, todos os *lifts* foram maiores que um.

3. Metodologia

Esta pesquisa está inserida no campo da Computação com aplicação na área da Educação. Utilizando a Ciência de Dados para realizar análises dos resultados de avaliações em larga escala, visando adquirir informações relevantes para a tomada de decisão. Por meio da aplicação de técnicas e métodos da Ciência de Dados, busca-se extrair percepções valiosas a partir dos resultados das avaliações em larga escala.

Em relação à forma da abordagem, trata-se de uma pesquisa quantitativa, visando realizar um estudo comparativo entre os achados de Lima (2023), que apresentou alguns indicativos de dificuldades dos estudantes do 9º ano do ensino fundamental de escolas públicas de Fortaleza na prova de Matemática do SPAECE para o ano de 2019; com os nossos resultados, no qual apresentamos as dificuldades identificadas para os estudantes de escolas públicas localizadas nas cidades do interior do Ceará. Para isso, utilizamos duas fases, na primeira aplicamos um processo da CD, e na segunda, realizamos uma análise comparativa.

3.1. Processo da CD

Para a obtenção dos resultados, seguimos a abordagem estruturada proposta por Cielien; Meysman e Ali (2016), que compreende seis etapas distintas, detalhadas a seguir:

1. Definir os Objetivos da Pesquisa: a primeira etapa desse processo é definir uma meta de pesquisa. O objetivo principal aqui é garantir que todas as partes interessadas entendam o quê, como e o porquê do projeto. Nesta pesquisa, temos como objetivo verificar se as dificuldades dos estudantes do 9º ano de escolas públicas da capital do Ceará (município de Fortaleza), identificadas por meio da análise dos dados da prova de Matemática do SPAECE, são similares às dos alunos que estudam em cidades do interior do Estado.

2. Coleta de Dados: a segunda etapa é a recuperação de dados. Esta etapa inclui encontrar dados adequados e obter acesso a eles. A coleta de dados para essa pesquisa foi feita por meio do acesso aos microdados não públicos do SPAECE referentes ao ano de 2019 fornecidos pela Célula de Informação, Indicadores Educacionais, Estudos e Pesquisas (CEIPE). Os dados cedidos trazem, principalmente, informações sobre os desempenhos dos alunos na prova de Português e Matemática do SPAECE. Esses dados estavam contidos em um arquivo do tipo Comma-Separated Values (CSV) com o tamanho de 317,7 Megabytes e 405.763 linhas de dados, no qual continham as informações das escolas, o detalhamento das respostas dos alunos em cada prova e as pontuações obtidas.

3. Preparação de Dados: na terceira etapa, agora com os dados brutos, é hora de prepará-los para o uso. Isso inclui realizar o tratamento e limpeza, modelando os dados conforme o modelo analítico que será utilizado. Nesta fase, foi realizada a verificação e o tratamento de valores ausentes ou inconsistentes nos dados. Dados faltantes podem comprometer a análise, portanto, foram removidos ou imputados de acordo com sua relevância.

Após a limpeza dos dados, aplicamos filtros para garantir o foco da pesquisa e a conformidade com os dados de Fortaleza obtidos por Lima (2023)). Nos dados, aplicamos um filtro para ter apenas as turmas do 9º ano e para manter apenas as questões de Matemática. As escolas estaduais foram removidas, restando apenas escolas municipais das cidades do interior do Ceará. Após a preparação dos dados, o tamanho do arquivo

diminuiu para 26,6 Megabytes e 77.662 linhas de dados.

4. Análise Exploratória de Dados: a quarta etapa tem o objetivo de obter uma compreensão aprofundada dos dados. Procuram-se padrões, correlações e desvios com base em técnicas visuais e descritivas. As percepções obtidas nesta fase permitiram uma modelagem eficiente. Nesta etapa, foi aplicado um script em linguagem de programação Python nos dados tratados, para a contabilização de descritores errados por cada aluno. Nesta etapa, analisamos quais foram os descritores que os alunos mais erraram e fizemos uma classificação ordenada pelos descritores com maior frequência de erros para melhor visualização e análise.

5. Modelagem de Dados: na quinta etapa realiza-se a busca para obter as percepções ou as previsões indicadas no termo de abertura do projeto. Com auxílio dos modelos analíticos ao diagnosticar seus resultados conforme os dados passados, é possível responder à questão de pesquisa após a compreensão dos dados. Nesta etapa submetemos os resultados da etapa anterior ao algoritmo FP-Growth para obter o modelo de dados. Para configuração do algoritmo utilizamos uma confiança de 80% e um suporte de 25%. A escolha do valor de suporte está menor do que foi utilizado no trabalho de Lima (2023) porque obtivemos poucos resultados quando utilizamos um suporte de 30%. Então, com o objetivo de ampliar a quantidade de resultados obtidos, adotamos um suporte mínimo de 25%.

6. Apresentação e Automação dos Resultados: a última etapa do modelo de CD, proposto por Cielen; Meysman e Ali (2016), é responsável pela apresentação dos resultados e automatização da análise, se necessário. Essa etapa será detalhada na seção de resultados.

3.2. Aspectos éticos da pesquisa

É importante considerar os aspectos éticos relacionados à proteção de dados pessoais, conforme a Lei Geral de Proteção de Dados Pessoais (LGPD). Os dados utilizados nesta pesquisa foram fornecidos pela SEDUC, por meio da CEIPE, e é fundamental respeitar as diretrizes estabelecidas pela legislação para garantir a privacidade e a segurança dos indivíduos envolvidos, que nesta pesquisa são os estudantes do 9º ano do ensino fundamental da rede pública do estado do Ceará. Nos microdados utilizados não é possível fazer nenhuma identificação de algum aluno específico, visto que não existe a exposição de nenhum dado pessoal, como nome ou número de identidade.

Além disso, outros dados utilizados nesta pesquisa foram devidamente anonimizados, conforme preconiza a LGPD, garantindo que não seja possível identificar individualmente os estudantes ou as instituições de ensino envolvidas, garantindo a confidencialidade e a privacidade das informações. Adicionalmente, todo o material coletado foi armazenado em repositório privado na nuvem, utilizando o serviço Google Drive, o que impede acessos não autorizados e reforça a segurança das informações tratadas. Essas medidas estão conforme as diretrizes estabelecidas pela LGPD para o tratamento de dados pessoais em pesquisas acadêmicas, conforme orientações da Autoridade Nacional de Proteção de Dados (ANPD) (ANPD, 2023).

4. Resultados e Discussões

Seguindo os procedimentos metodológicos definidos na seção anterior, com a aplicação do processo da CD, definido por Cielen; Meysman e Ali (2016), obtivemos inicialmente a Tabela 1, que apresenta os resultados da frequência dos erros nos descritores de Matemática dos municípios do interior do Ceará (obtidos por esta pesquisa), juntamente com as frequências obtidas por Lima (2023) para o município de

Fortaleza. Como forma de facilitar o entendimento, o Quadro 1 apresenta o significado de todos os descritores do SPAECE utilizados nesta pesquisa.

Observe que a Tabela 1 traz os cinco erros mais frequentes cometidos pelos alunos do interior do Ceará e de Fortaleza, conforme os resultados do 9º ano na prova do SPAECE 2019, em Matemática. Ao analisarmos os resultados, podemos perceber que os descritores D067, D007, D051 e D050 são os mais problemáticos, tanto no interior quanto para a capital Fortaleza, ocupando as primeiras posições nos dois resultados. No entanto, enquanto no interior o descritor D065 aparece na 5ª posição, em Fortaleza temos o descritor D015.

Tabela 1. Frequência dos cinco maiores erros nos descritores do SPAECE 2019, no 9º ano do ensino fundamental em Matemática

Resultados	Descritores	Frequência
Essa pesquisa - escolas do interior do Ceará	D067	52672
	D007	48414
	D051	44388
	D050	42025
	D065	41964
Lima (2023) - escolas da capital do Ceará	D067	13021
	D007	10905
	D051	10557
	D050	10375
	D015	10330

Quadro 1. Significados dos descritores utilizados do SPAECE

Descritor	Descrição/Significado
D007	Resolver situação problema utilizando mínimo múltiplo comum ou máximo divisor comum com números naturais
D011	Ordenar ou identificar a localização de números racionais na reta numérica
D013	Reconhecer diferentes representações de um mesmo número racional, em situação-problema
D015	Resolva o problema utilizando adição ou subtração com números racionais representados na forma fracionária (mesmo denominador ou denominadores diferentes) ou na forma decimal
D017	Resolver situação problema utilizando porcentagem
D021	Efetuar cálculos com números irracionais, utilizando suas propriedades
D024	Fatorar e simplificar expressões algébricas
D049	Resolver problema envolvendo semelhança de figuras planas
D050	Resolver situação problema aplicando o Teorema de Pitágoras ou as demais relações considerações no triângulo retângulo
D051	Resolver problemas usando as propriedades dos polígonos (soma dos ângulos internos, número de diagonal e cálculo do ângulo interno de polígonos regulares)
D065	Calcular o perímetro de figuras planas, numa situação problema
D067	Resolver problema envolvendo o cálculo de área de figuras planas
D069	Resolver problema envolvendo noções de volume

Ao observarmos a Tabela 1, é possível perceber que os quatro principais erros

em descritores (tanto para o interior, quanto para a capital), concentram-se no tema II do SPAECE, “Convivendo com a Geometria”, onde estão os descritores D050 e D051. Analisando especificamente os dados do interior, há também uma concentração no tema III, “Vivenciando as Medidas”, com os descritores D067 e D065.

A capital Fortaleza, por sua vez, traz a presença dos descritores D007 e D015, que fazem parte do tema I, “Interagindo com Números e Funções”. Portanto, ao observar apenas os temas dos descritores errados com mais frequência, os dados indicam que as fragilidades da capital estão mais relacionadas a conceitos matemáticos básicos, enquanto no interior elas se voltam a conceitos geométricos e de mensuração. Essa constatação é reforçada quando verificamos que o descritor D015, que aparece entre os cinco maiores problemas em Fortaleza, encontra-se na 21ª posição na classificação do interior. Os dados sugerem que os estudantes de Fortaleza possuem maior dificuldade em assuntos mais básicos, como contas com números racionais, enquanto os do interior apresentam dificuldades em assuntos mais complexos envolvendo geometria.

Embora a utilização de frequências nos erros dos descritores seja amplamente adotada no Brasil, esse tipo de análise não consegue oferecer uma melhor compreensão das dificuldades, portanto, outras abordagens devem ser utilizadas, como a aplicação do algoritmo FP-Growth (Lima, 2023).

Como forma de trazer mais elementos sobre as dificuldades em Matemática, a Tabela 2 lista as associações, por meio da aplicação do algoritmo FP-Growth, entre os erros nos descritores de Matemática nos municípios do interior do Ceará e as associações obtidas por Lima (2023) para a capital Fortaleza. Nela são apresentadas as cinco associações mais relevantes entre os erros nos descritores, tanto para os municípios do interior como para a capital Fortaleza. Essas relações foram classificadas com base no suporte (frequência da ocorrência no conjunto de dados) e na confiança (probabilidade de o consequente ocorrer dado o antecedente).

Tabela 2. Associações mais frequentes entre erros nos descritores do SPAECE 2019, para o 9º ano do ensino fundamental em Matemática

Resultados	Antecedente/s	Consequente/s	Confiança (%)	Suporte (%)
Escolas municipais do interior do Ceará	D015	D067	86,78	38,19
	D013 e D069	D021	82,26	31,71
	D069 e D021	D013	80,09	31,71
	D048	D007	81,98	30,33
	D017 e D051	D067	82,80	27,52
Escolas municipais da capital do Ceará (Lima, 2023)	D021	D013	83,65	48,81
	D021	D069	81,99	47,84
	D015	D067	89,47	46,19
	D011	D067	81,73	42,71
	D049	D024	81,66	42,69

Na Tabela 2, as cinco associações mais frequentes de Fortaleza, nos ajudam a montar o quadro dos principais problemas da cidade e analisar suas causas e efeitos devido às associações de antecedente e consequente. O problema mais frequente de Fortaleza (apresentado na Tabela 1) foi o descritor D067. Esse descritor está presente nas associações geradas pelo FP-Growth (apresentadas na Tabela 2) com dois antecedentes, o descritor D015, com 89% de confiança e 46% de suporte; e o descritor D011, com 81% de confiança e 42% de suporte.

Esses resultados nos auxiliam a entender que grande parte das dificuldades

com as figuras planas é resultado da falta de habilidades dos alunos com os números racionais. A pesquisa de Chaves e Ramirez (2025), que analisou os dados do Sistema de Avaliação Educacional de Fortaleza (SAEF) para o ensino fundamental no ano de 2019, também encontrou resultados semelhantes. Eles observaram que os estudantes tinham dificuldades com o descritor D13 do SAEF (Reconhecer diferentes representações de um mesmo número racional, situação-problema), tendo a maior taxa de erros. Embora as pesquisas citadas tenham analisado diferentes provas em larga escala, seus resultados estão alinhados.

A Tabela 2 também apresenta a relação entre o descritor D021 e o D013, com o maior suporte apresentado de 48,8%, apontando para dificuldades dos alunos em transitar dos estudos dos números irracionais para os racionais. Complementados com as associações envolvendo o descritor D021 e D069 e a relação do descritor D049 e D024, que reforçam a presença de dificuldades em tópicos de geometria como volume e figuras planas, quanto em álgebra, corroborando com o trabalho de Pontes (2023), que identificou dificuldades dos alunos de Fortaleza em geometria, analisando os dados do SAEB.

Quando analisamos as cinco principais associações do interior, podemos perceber que a associação com maior suporte tem como consequente o descritor D067 que é o problema mais frequente do interior (de acordo com a Tabela 1), mas, com os resultados obtidos percebemos o D015 como antecedente, e também percebemos o D017 e D051 como antecedentes do D067, demonstrando que no interior, além da dificuldade com os números racionais, os conceitos percentuais e poligonais também estão afetando o desempenho em geometria. Resultado que coincide com os achados do estudo de Santos (2010), no qual ela analisa os resultados da aplicação do SPAECE no ano de 2008 em todo o Ceará para o ensino médio, em que a presença do descritor D017 e D051 aparecem com uma taxa de acerto menor do que 40% no 1º ano do ensino médio.

Além disso, a associação entre o D013 e D069 como antecedente do D021 com 80,09% de confiança, indica que as dificuldades nas noções de volume e compreensão dos números racionais afetam o aprendizado dos números irracionais. De acordo com o estudo de Santos (2010), percebemos que esses descritores não passam de 30% de acerto no 1º ano do ensino médio. A associação entre D048 e D007 sugere uma carência tanto em geometria básica quanto em operações fundamentais de divisibilidade no interior. Isso ratifica a pesquisa de Silva e Lima (2013), que por meio de entrevistas com professores de escolas de Iguatu, do interior do Ceará, identificaram que a maior dificuldade dos alunos do 6º ao 9º ano está nas operações básicas.

Quando comparamos os dois resultados (interior e capital), temos dificuldades semelhantes em alguns assuntos, como na relação do descritor D067 com o D015, reforçando como os números racionais afetam significativamente a resolução de problemas geométricos e que se tornaram um problema para o Ceará como um todo. Também foram observadas associações envolvendo os descritores D013, D069 e o D021, aparecendo em ambos os resultados, evidenciando dificuldades na compreensão dos números racionais e irracionais e em conceitos volumétricos.

Com a comparação dos resultados, também percebemos algumas diferenças. Em Fortaleza, o descritor D011 aparece como antecedente do D067, demonstrando maior dificuldade dos estudantes com a reta numérica e números racionais. Já nas escolas do interior, como antecedentes do descritor D067, temos os descritores D017 e D051, indicando que os alunos do interior possuem dificuldade com cálculos envolvendo porcentagem e o entendimento de conceitos geométricos mais básicos. Além disso, apenas no interior os descritores D048 e D007 aparecem entre os 5 principais, já em Fortaleza são o D049 e D024.

A partir dos resultados da análise, podemos discutir o nível e a qualidade do ensino em Fortaleza e no interior do estado do Ceará, considerando os padrões de erro identificados nas associações dos descritores do SPAECE 2019 pela saída do FP-Growth (disponíveis na Tabela 2). Os dados sugerem que há diferenças na qualidade do ensino de Matemática das duas regiões. Como apresentado nos resultados, podemos perceber que as dificuldades de Fortaleza procedem de assuntos mais complexos e abstratos como o D021, D011 e o D015, dando a entender que Fortaleza possui uma boa base sobre os conceitos matemáticos e tem dificuldade em conteúdos mais avançados de geometria, números racionais e irracionais. Por outro lado, o Interior apresenta dificuldades em conceitos matemáticos mais elementares como o D007, D048, D017 e D051, indicando que a qualidade do ensino no Interior está a um nível mais baixo em comparação à capital Fortaleza.

Observando os resultados, é perceptível a importância da aplicação de uma técnica mais elaborada para a análise dos dados, que permite uma maior precisão nos resultados. Quando olhamos apenas para a frequência dos descritores mais errados, era possível concluir que o interior, mesmo que por pouco, estava em uma situação melhor que Fortaleza em matemática, mas ampliando e aprofundando a análise com o algoritmo do FP-Growth, é possível concluir pelos resultados que, na realidade, Fortaleza se encontra melhor em matemática quando comparamos com o interior. Essa análise mais precisa dos dados acaba auxiliando em tomadas de decisões educacionais mais sólidas e condizentes com a realidade.

5. Conclusões

Neste estudo utilizamos os microdados do SPAECE 2019 para realizar uma análise comparativa dos resultados do 9º em Matemática entre os municípios do Interior do Ceará com os resultados obtidos por Lima (2023), que analisou os resultados de Matemática no município de Fortaleza. Buscamos investigar se as dificuldades apresentadas em Fortaleza seriam as mesmas dificuldades observadas nos municípios do Interior, de forma que percepções valiosas pudessem ser identificadas para possíveis melhorias na qualidade da educação pública.

Para a investigação dos resultados foi aplicado o processo da CD descrito por Cielén; Meysman e Ali (2016). Com os dados da prova do SPAECE, foi aplicada a técnica da CD utilizando o algoritmo FP-Growth, além de serem construídas tabelas com o método de ranqueamento dos erros pela frequência para que, enfim, a análise comparativa entre as pesquisas ocorresse.

Ao final desse estudo, podemos concluir que os resultados apontam semelhanças e diferenças entre as dificuldades de cada região, destacando as dificuldades semelhantes com números racionais e geometria, mas, ao mesmo tempo, evidenciando que em Fortaleza há uma necessidade de melhoria em conceitos matemáticos mais avançados, enquanto no Interior se destaca a dificuldade em assuntos matemáticos elementares. Demonstrando que as melhorias nas políticas e práticas pedagógicas devem ser específicas a cada região. Os achados podem também ser utilizados para melhorar a forma como as formações de professores são planejadas. Buscando, assim, atender aos anseios e necessidades dos docentes, tanto da capital como do interior do Ceará.

Embora o estudo tenha fornecido percepções importantes sobre as dificuldades em Matemática dos alunos de Fortaleza e do interior do 9º ano, ele possui algumas fragilidades, como foco apenas nos resultados do 9º pela prova do SPAECE, além do uso exclusivo do FP-Growth e a utilização apenas dos dados do ano de 2019.

Como trabalhos futuros, sugerimos expandir os anos avaliados para a construção de uma série histórica, de forma que tenhamos um maior detalhamento das dificuldades dos estudantes ao longo dos anos, não apenas para a disciplina de matemática, mas também para português. Também é interessante inserir outras técnicas da CD que podem trazer novos elementos para a análise.

Referências

- ANPD, A. N. de Proteção de D. **Guia orientativo: tratamento de dados pessoais para fins acadêmicos e para a realização de estudos e pesquisas**. 2023. <<https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/documentos-de-publicacoes/web-guia-anpd-tratamento-de-dados-para-fins-academicos.pdf>>. Acesso em: 18 set. 2024.
- Bauer, A.; Gatti, B. A.; Tavares, M. R. **Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origem e pressupostos**. Florianópolis: Insular, 2013.
- Bonamino, A.; Sousa, S. Z. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo na escola. **Educação e Pesquisa**, São Paulo, v. 38, n. 2, p. 373–388, 2012.
- Bérgamo, L.; Oliveira, V. C. de. Big data em nuvem: uma pesquisa bibliográfica e comparativa entre bancos de dados e servidores locais e cloud. **SITEFA**, v. 2, n. 1, p. 404–411, 2019.
- Ceará, G. do Estado do. **97% das crianças cearenses estão alfabetizadas**. 2024. <<https://www.ceara.gov.br/2024/05/27/spaace-2023-97-das-criancas-cearenses-estao-alfabetizadas/>>. Acesso em: 01 out. 2024.
- Ceará, S. da Educação do Estado do. **SPAECE**. 2020. <<https://www.seduc.ce.gov.br/spaace/>>. Acesso em: 18 set. 2024.
- Chaves, M. M.; Ramirez, A. R. G. Metodologias ativas no ensino da matemática: uma análise das avaliações diagnósticas de matemática da rede municipal de ensino de fortaleza e a metodologia de rotação por estações. **Caderno Pedagógico**, v. 22, n. 5, p. 1–17, 2025.
- Cielen, D.; Meysman, A. D. B.; Ali, M. **Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools**. Shelter Island: Manning, 2016.
- Gatti, B. A. Testes e avaliações do ensino no Brasil. **Educação e Seleção**, v. 16, p. 33–42, 2013. Disponível em: <<https://publicacoes.fcc.org.br/edusel/article/view/2623>>.
- Grossman, M. **Education and Nonmarket Outcomes**. Cambridge, MA, 2005. Acessado em 27 set. 2024. Disponível em: <<http://www.nber.org/papers/w11582>>.
- Klein, R.; Fontanive, N. S. Avaliação em larga escala: uma proposta inovadora. **Em Aberto**, Inep, v. 15, n. 66, 1995.
- Lima, M. V. D. A. **(Re)estruturação do modelo de planejamento das ações de formação continuada em serviço de professores na rede municipal de Fortaleza: novas possibilidades por meio da ciência de dados educacionais**. Tese (Doutorado) — Universidade Estadual do Ceará, Fortaleza, 2023. Acesso em: 02 set. 2024.
- Lorenzetti, C. D. C.; Telöcken, A. V. Estudo comparativo entre os algoritmos de mineração de dados random forest e j48 na tomada de decisão. In: Universidade de Cruz Alta (UNICRUZ). **Simpósio de Pesquisa e Desenvolvimento em Computação**. Cruz Alta, RS, 2016.
- Mariano, M. A. **Comparação de algoritmos paralelos para a extração de regras de associação no modelo de memória distribuída**. Dissertação (Mestrado) — Universidade Federal de Mato Grosso do Sul (UFMS), Campo Grande, MS, nov 2011. Acesso em: 22 set. 2024. Disponível em: <<https://repositorio.ufms.br/handle/123456789/1674>>.

- Pontes, M. M. de. Um estudo comparativo da análise dos erros nos descritores do spaee e do saeb com uso do algoritmo fp-growth. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) — Universidade Federal do Ceará, Campus de Russas. 2023.
- Prado, T. A. de Knezt López de; Siqueira, G. M.; Júnior, W. M.; Carvalho, M. L. B. de. pifp-growth: Um algoritmo paralelo para geração incremental de regras de associação. In: **Anais do Simpósio em Sistemas Computacionais de Alto Desempenho (WSCAD)**. Vitória: Sociedade Brasileira de Computação, 2002. p. 76–83.
- Saldanha, R. de F.; Barcellos, C.; Pedroso, M. de M. Ciência de dados e big data: o que isso significa para estudos populacionais e da saúde? **Cadernos de Saúde Coletiva**, v. 29, n. esp., p. 51–58, 2021.
- Santos, F. D. G. dos. **Impactos gerados pelo Sistema Permanente de Avaliação da Educação Básica do Estado do Ceará (SPAEECE) na melhoria do ensino e aprendizagem no Ensino Médio**. Dissertação (Dissertação (Mestrado em Educação)) — Universidade Federal do Ceará (UFC), Fortaleza, CE, 2010. Acesso em: 22 set. 2024. Disponível em: <http://www.repositorio.ufc.br/handle/riufc/3089>.
- Silva, A. F. da; Lima, F. J. de. Educação e currículo: algumas reflexões sobre a aversão de alunos pela aprendizagem de conteúdos matemáticos. In: **Anais do VII Congresso Iberoamericano de Educação Matemática (CIBEM)**. Funes: Repositório de Documentos de Educação Matemática, 2013. Acesso em: 5 out. 2024. Disponível em: <https://funes.uniandes.edu.co/funes-documentos/educacao-e-curriculo-algumas-reflexoes-sobre-a-aversao-de-alunos-pela-aprendizagem-de-conteudos-matematicos/>.
- Silva, L. A. da; Peres, S. M.; Boscardioli, C. **Introdução à mineração de dados: com aplicações em R**. 1. ed. Rio de Janeiro: Elsevier, 2016.
- Werle, F. O. C. Políticas de avaliação em larga escala na educação básica: do controle de resultados à intervenção nos processos de operacionalização do ensino. **Ensaio: Avaliação e Políticas Públicas em Educação**, Fundação CESGRANRIO, Rio de Janeiro, v. 19, n. 73, p. 769–792, 2011.