ESCOLA DE MATEMÁTICA
MESTRADO PROFISSIONAL

FÁBIO GONÇALVES DE OLIVEIRA

AGRUPAMENTO INTELIGENTE DE ESTUDANTES PARA INTERVENÇÕES PEDAGÓGICAS: REUTILIZANDO DADOS ADMINISTRATIVOS NA GESTÃO DA APRENDIZAGEM MATEMÁTICA

RIO DE JANEIRO

2025

Agrupamento inteligente de estudantes para intervenções pedagógicas: reutilizando dados administrativos na gestão da aprendizagem matemática

Por

Fábio Gonçalves de Oliveira

Trabalho de conclusão do curso de Pós-Graduação Stricto Sensu, Mestrado Profissional em Matemática em Rede Nacional, para aprimoramento da formação profissional de professores da educação básica pela Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre.

Orientador: Alexandre Sousa da Silva

Rio de Janeiro, 24 de março de 2025

Catalogação informatizada pelo autor:

O48 Oliveira, Fábio Gonçalves de
Agrupamento inteligente de estudantes para intervenções pedagógicas: reutilizando dados administrativos na gestão da aprendizagem matemática / Fábio Gonçalves de Oliveira. -- Rio de Janeiro : UNIRIO, 2025.
60f

Orientador: Alexandre Sousa da Silva.
Dissertação (Mestrado) - Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Matemática, 2025.

1. Dados administrativos. 2. Gestão escolar por dados e evidências. 3. Desempenho Escolar em Matemática. I. Silva, Alexandre Sousa da, orient. II. Título.

FÁBIO GONÇALVES DE OLIVEIRA

AGRUPAMENTO INTELIGENTE DE ESTUDANTES PARA INTERVENÇÕES PEDAGÓGICAS: REUTILIZANDO DADOS ADMINISTRATIVOS NA GESTÃO DA APRENDIZAGEM MATEMÁTICA

Trabalho de conclusão do curso de Pós-Graduação Stricto Sensu, Mestrado Profissional em Matemática em Rede Nacional, para aprimoramento da formação profissional de professores da educação básica pela Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre.

Aprovado em 18 de março de 2025

Membros da Banca:

gov.br Documento assinado digitalmente
ALEXANDRE SOUSA DA SILVA
Data: 15/04/2025 10:25:10-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Alexandre Sousa da Silva (Orientador – Unirio)

gov.br Documento assinado digitalmente
MICHEL CAMBRAINHA DE PAULA
Data: 15/04/2025 12:15:17-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Michel Cambrainha de Paula (Unirio)

gov.br Documento assinado digitalmente
JONY ARRAIS PINTO JUNIOR
Data: 22/04/2025 10:05:07-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Jony Arrais Pinto Júnior (Uff)

RESUMO

O uso de dados administrativos é amplamente reconhecido como fundamental para embasar políticas públicas orientadas por evidências. Entretanto, no cotidiano escolar esses dados costumam ser subaproveitados, dificultando sua conversão em informações úteis para aprimorar a gestão e a qualidade do ensino. Este estudo examina como avaliações padronizadas e outros registros administrativos podem ser empregados de modo mais efetivo na identificação de lacunas de aprendizagem, no agrupamento estratégico de estudantes e na adequação de práticas pedagógicas, visando otimizar o uso dos recursos educacionais. A literatura indica que estratégias de agrupamento podem beneficiar tanto alunos com dificuldades quanto aqueles com desempenho superior, promovendo um ensino mais equitativo e eficiente. Contudo, as práticas de agrupamento são geralmente informais e raramente sistematizadas, o que limita seu potencial transformador. O resultado aqui apresentado consiste em um sistema — desenvolvido com Cursor AI em interação com o ChatGPT-4o — que, utilizando dados brutos de avaliações censitárias, possibilita o agrupamento de estudantes, a criação de mapas de assentos, planos de estudo individualizados e novas visualizações de dados. Além disso, oferece relatórios de desempenho em análises transversais e longitudinais, tanto em nível coletivo quanto individual. Espera-se, assim, auxiliar na tomada de decisões e agilizar iniciativas que exijam diferentes formas de agrupamento, contribuindo para a qualidade e a equidade educacional.

Palavras-chave: Dados administrativos; Gestão escolar por dados; Agrupamento de estudantes; Desempenho Matemático; Ferramentas pedagógicas digitais; Decisão por evidências

*A*BSTRACT

The use of administrative data is widely acknowledged as a key element in informing evidence-based public policies. Nonetheless, within the daily routines of schools, such data are frequently underutilized, limiting their potential to generate actionable insights for improving educational management and instructional quality. This study explores how standardized assessments and other administrative records can be more effectively leveraged to identify learning gaps, strategically group students, and tailor pedagogical practices with the goal of optimizing educational resource allocation. The literature suggests that student grouping strategies can benefit both low- and high-achieving learners, promoting more equitable and efficient instruction. However, these practices are often informal and lack systematic implementation, which constrains their transformative impact. The present work introduces a system—developed using Cursor AI in conjunction with ChatGPT-4o—that processes raw data from large-scale assessments to enable student grouping, the generation of seating charts, individualized learning plans, and innovative data visualizations. Additionally, it offers performance reports through both cross-sectional and longitudinal analyses at individual and collective levels. The system aims to support data-informed decision-making and streamline educational initiatives requiring differentiated groupings, ultimately contributing to enhanced equity and quality in education.

Keywords: Administrative data; Data-driven school management; Student grouping; Mathematics achievement; Digital pedagogical tools; Evidence-based decision-making.

Aproximação ao tema 8
1 - Introdução 10
2 - Revisão da Literatura 12

2.1 - O aproveitamento de dados administrativos na educação 12
2.2 - As provas padronizadas como fontes de dados 13
2.3 - Bases de dados para análises transversais e longitudinais 13
2.4 - As novas ferramentas de análise 14
2.5 - Estratégias de agrupamento de estudantes 14

3 - Aspectos Metodológicos 19

3.1 - O caso da Prefeitura do Rio de Janeiro 19
3.2 - A estrutura da base de dados 20
3.3 - Especificação da Funcionalidade Mínima do Produto 24
3.4 - Arquitetura e escolhas de implementação 26

4 - Resultados: O Produto Educacional SIGGA 30

4.1 - Obtenção dos Dados 30
4.2 - Enturmação por Nível de Habilidade 32
4.3 - Criação de Mapas de Assentos 33
4.4 - Rotinas de Estudo Personalizadas. 35
4.5 - Análises Longitudinais 35

5 - Discussão 38

5.1 - Sugestões para Estudos Futuros 39
5.2 - Limitações do estudo 40
5.3 - Considerações Finais 42

Referências 43
Anexo 1 - Rotinas de Estudo Personalizadas – Matemática 45
Anexo 2 - Prompts selecionados para a geração da versão MVP do app 47
Anexo 3 - Formato do Arquivo de Reenturmação 49
Anexo 4 - Formato do Arquivo Mapa de Assentos 50
Anexo 5 - Formato do Arquivo Balanço Anual 51

Anexo 6 - Exploração preliminar do Vínculo entre Literacia e Numeracia 52
Anexo 7 - Sistema de Categorização das Notas do SIGGA 55
Anexo 8 - Manual do Usuário SIGGA 58

Aproximação ao tema

Por cerca de duas décadas, minha principal atividade profissional foi a gestão de equipes. Atuei no setor privado, posteriormente em uma ONG[1] voltada à educação digital e, após concluir um mestrado em administração pública, assumi uma diretoria do Instituto de Planejamento da Prefeitura do Rio de Janeiro. Somente depois dessas experiências cursei a licenciatura que me permitiu atuar a partir de 2018 como professor de matemática em escolas públicas.

Se, por um lado, tive a sorte de ingressar em uma escola cujo diretor era um entusiasta de práticas inovadoras, por outro me surpreendeu a autonomia do trabalho docente, que coexistia com uma notável falta de supervisão e apoio das estruturas centrais às atividades cotidianas do professor.

Durante o longo hiato imposto pela pandemia, eu lecionava remotamente para no máximo 20 dos cerca de 250 alunos matriculados em minhas turmas. O severo impacto que esse período teria sobre as crianças já era evidente. Quando finalmente retornamos às salas de aula escolares, em março de 2021, deparamo-nos com um cenário já conhecido, mas agora agravado: altos índices de analfabetismo funcional entre os estudantes do 6º ano, sendo que aproximadamente 20% demonstravam total incapacidade de leitura, segundo avaliação realizada pelos professores de língua portuguesa.

Que impacto essas deficiências teriam na aprendizagem matemática? Inicialmente, a hipótese central do estudo supunha uma conexão deletéria entre o uso ineficaz da língua nativa e o desempenho em matemática, e o produto pretendia justamente identificar os estudantes que enfrentam essa dificuldade adicional. Porém, aprofundar essa linha de pesquisa entraria em conflito com o escopo de um mestrado profissional, seja pelo tempo de campo necessário, seja pelas implicações éticas — como identificar alunos com atraso sem oferecer-lhes os meios de superá-lo.

[1] Comitê para a Democratização da Informática, organização não governamental que levava computadores a comunidades no Brasil e no exterior, oferecendo capacitação em cidadania e informática – como se dizia à época.

Por essas razões, o escopo do produto educacional foi ajustado para evitar questões que exigissem consentimento formal, focando-se no desenvolvimento de uma ferramenta de agrupamento inteligente de estudantes e visualização aprofundada dos dados brutos já disponíveis, mas subutilizados. Apenas os próprios praticantes, que já possuem acesso aos dados, receberão as listagens nominais e visualizações que individualizam desempenhos.

1 - Introdução

O ensino fundamental é uma etapa crucial no desenvolvimento das competências de letramento e numeracia, que constituem a base para o progresso educacional e profissional dos estudantes (BANERJEE et al, 2007; ALVES; SOARES, 2007).

O agrupamento de estudantes, que inclui a enturmação como uma de suas modalidades, é uma atividade recorrente tanto na gestão escolar quanto na gestão da aprendizagem. Esse processo não se limita às enturmações no início de cada ano letivo, mas também atende a demandas como transferências, seleção de bolsas de estudo, indicação para aulas de reforço, atividades extracurriculares e, no dia a dia da sala de aula, para a realização de provas, trabalhos e dinâmicas diversas.

Na prática, o agrupamento de estudantes — incluindo a enturmação — costuma basear-se principalmente no conhecimento tácito de professores e gestores, exigindo grande esforço das equipes escolares e impactando diretamente a dinâmica de ensino e a organização da sala de aula. Entretanto, quando realizado com critérios bem definidos, o agrupamento contribui para a redução de desigualdades educacionais e para o uso mais eficiente de recursos pedagógicos (ALVES; SOARES, 2007; BETTS; SHKOLNIK, 2000).

Entre os vários critérios disponíveis para o agrupamento de estudantes, o emprego do nível de habilidade se destaca por permitir tanto a identificação de alunos com defasagens específicas quanto a promoção de talentos, muitas vezes subestimados em sistemas que priorizam o “aluno médio” (LECHUGA; DOROUDI, 2023). Essa estratégia de enturmação torna possível adaptar planos de ensino e rotinas de estudo com maior precisão, contribuindo para um ambiente educacional mais inclusivo e eficiente (ALVES; SOARES, 2007; BETTS; SHKOLNIK, 2000).

Este estudo apresenta uma ferramenta que integra dados administrativos — obtidos tanto de avaliações padronizadas quanto de registros rotineiros da escola — para agrupar estudantes de modo mais sistemático e fundamentado. Embora esses dados constituam um repositório de grande valor (FIGLIO; KARBOWNIK; SALVANES, 2017), costumam encontrar-se dispersos em diferentes plataformas, dificultando sua interpretação no dia a dia escolar. Para sanar esse problema, o sistema aqui proposto consolida essas informações e automatiza a geração de relatórios e visualizações, oferecendo subsídios para gestores e professores. Além disso, recorre-se à

inteligência artificial a fim de produzir relatórios personalizados sobre o desempenho dos alunos.

Ao combinar dados administrativos, algoritmos de agrupamento inteligente e ferramentas computacionais, busca-se não apenas agilizar o trabalho da gestão escolar na alocação de estudantes em turmas, mas também facilitar a construção de grupos homogêneos ou heterogêneos, em consonância com as necessidades pedagógicas dos professores. Dessa forma, torna-se possível, por exemplo, organizar assentos em sala, elaborar feedbacks personalizados e acompanhar o desempenho de cada turma ao longo do tempo.

Ademais, a abordagem proposta extrapola o ambiente da sala de aula ou mesmo o escopo de uma única escola. Ela pode ser estendida para subsidiar as Coordenadorias Regionais e a Secretaria de Educação na formulação de políticas públicas mais eficazes (GATTI, 2004), ao viabilizar a identificação objetiva de necessidades e potencialidades dos estudantes. Esse uso estruturado dos dados administrativos permite, por exemplo, estimar quantos alunos necessitam de reforço em toda a rede e localizar talentos para programas mais avançados, evidenciando a relevância da gestão informada por dados como via para promover equidade e qualidade na educação.

Diante desse cenário, esta dissertação tem por objetivos:

1. Desenvolver uma ferramenta que utilize avaliações padronizadas e outros registros rotineiros da escola para automatizar e otimizar o agrupamento de estudantes de forma sistemática e fundamentada.
2. Apoiar intervenções pedagógicas, a partir de dados já coletados, auxiliando professores e gestores na alocação de recursos pedagógicos, na elaboração de relatórios personalizados e no planejamento de aulas adaptadas ao perfil de cada turma, contribuindo para uma gestão escolar mais equânime e eficaz.

2 - Revisão da Literatura

Esta seção apresenta uma breve revisão da literatura, abordando dois tópicos: a importância da análise de dados administrativos e o papel da tecnologia na avaliação e melhoria da educação. Esses tópicos fornecem o contexto para a discussão que encerra a seção, um levantamento de estratégias de agrupamento de estudantes para diferentes finalidades — tema essencial do produto educacional desenvolvido.

2.1 - O aproveitamento de dados administrativos na educação

Os dados administrativos são aqueles coletados rotineiramente pelas instituições públicas (exemplos incluem registros de saúde, educação, segurança pública, serviços sociais) e constituem uma ferramenta importante para a formulação de políticas públicas. Figlio et al. (2017) destacam que a crescente capacidade de armazenamento de dados e computação tornaram possível coletar, gerenciar e analisar dados em estratégias e escala inconcebíveis há poucos anos. Esses dados, coletados inicialmente para fins administrativos, oferecem oportunidades notáveis para expandir nosso conhecimento sobre comportamentos e resultados ao longo da vida dos indivíduos (FIGLIO et al., 2017).

Esse potencial de uso dos dados administrativos se insinua em estudos que empregam dados públicos do Instituto Brasileiro de Geografia e Estatística (IBGE) e do Instituto Nacional de Estudos e Pesquisas Educacionais (Inep). Algumas pesquisas analisam esses dados de forma inovadora, evidenciando disparidades regionais e socioeconômicas e orientando políticas mitigatórias (ALVES, 2020).

No entanto, essas bases públicas não se comparam às bases administrativas internas que, embora pouco exploradas, oferecem um nível de detalhamento muito maior — inclusive viabilizando o acompanhamento longitudinal de cada indivíduo, permitindo estabelecer cadeias causais (RØED; RAAUM, 2003). Nesse sentido, o uso de dados administrativos, por sua natureza mais granular, desponta como um caminho promissor para estudos futuros, permitindo aprofundar o entendimento das realidades educacionais e aproveitar ao máximo o *framework* de pesquisas já existentes.

2.2 - As provas padronizadas como fontes de dados

As provas padronizadas desempenham um papel crucial na avaliação e monitoramento do desempenho dos alunos de maneira sistemática e comparável. Bassetto (2021) utilizou dados do SARESP em conjunto com questionários estruturados para analisar disparidades no desempenho dos estudantes, destacando como esses exames fornecem uma base consistente de dados para análises longitudinais e comparativas.

Além disso, a disciplina de Mineração de Dados Educacionais (EDM no original em inglês) pode ser aplicada aos resultados de provas padronizadas para identificar padrões e prever resultados futuros (FONSECA; NAMEN, 2016). Embora ainda incipiente no Brasil, a EDM pode ajudar a personalizar a educação e a desenvolver intervenções mais eficazes para melhorar o desempenho dos alunos, especialmente no caso da EaD (BAKER et al., 2011).

2.3 - Bases de dados para análises transversais e longitudinais

A análise longitudinal de dados educacionais permite uma compreensão mais profunda do desenvolvimento dos alunos ao longo do tempo. Para Figlio et al. (2017), a coleta contínua de dados administrativos possibilita o estudo de questões educacionais de uma maneira que antes não era possível, permitindo a identificação de tendências e padrões que podem informar intervenções mais eficazes.

Mais comum que a análise longitudinal, a análise transversal ou segmentada também oferece valiosas informações. De Cristo e Canan (2022) agrupam dados sobre professores das IFs por regiões e comparam variáveis como qualificação, regime de trabalho e vínculo com a administração pública, permitindo identificar disparidades regionais e necessidades específicas de cada segmento. Essa abordagem segmentada pode revelar diferenças importantes no desempenho e nas condições de ensino entre diferentes grupos de alunos ou regiões, informando políticas mais específicas e direcionadas.

Rodrigues et al. (2024) mapeiam algoritmos de IA para prever a evasão universitária, utilizando principalmente registros de notas, presença, dados socioeconômicos e avaliações de satisfação destacando a importância de segmentar os estudantes com base em suas características e comportamentos de risco. Valendo-se desta abordagem transversal é possível desenvolver intervenções mais precisas e direcionadas, como programas de tutoria para alunos em risco de evasão ou reforço

acadêmico para aqueles com dificuldades persistentes visando melhorar a retenção dos alunos.

2.4 - As novas ferramentas de análise

Atualmente é possível encontrar na literatura artigos que tratam do uso crescente de Inteligência artificial para análise de dados educacionais. As técnicas de aprendizado de máquina, incluindo redes neurais, estão sendo cada vez mais aplicadas na educação para prever comportamentos e resultados educacionais. Rodrigues et al. (2024) discutem o uso de algoritmos de IA para prever a evasão universitária, destacando a importância dessas ferramentas para identificar padrões complexos nos dados educacionais. Já Hartmann et al. (2017) focam na aplicação de tecnologias para avaliar a complexidade de textos não ficcionais. Eles desenvolvem um sistema de classificação automática que determina o nível de complexidade dos textos, destacando como as ferramentas tecnológicas podem ser utilizadas para garantir que os materiais educativos sejam adequados ao nível de habilidade dos estudantes.

2.5 - Estratégias de agrupamento de estudantes

Informações como resultados de avaliação em larga escala podem ser utilizadas no ambiente educacional para favorecer as práticas pedagógicas. Essas informações, quando adequadamente integradas à preparação de aula, permitem intervenções direcionadas que atendem tanto às deficiências quanto às potencialidades de cada estudante (LOUREIRO; RODRIGUES, 2024).

Serão apresentadas, a seguir, diferentes estratégias de agrupamento de estudantes usando exclusivamente o nível de habilidade. Para subsidiar esses processos, propõe-se a utilização de resultados de provas de nivelamento ou de avaliações padronizadas. Com a integração de outras fontes de dados administrativos, como registros disciplinares e avaliações socioemocionais, estes critérios podem e devem ser expandidos. Defende-se aqui que as estratégias de agrupamento podem contribuir para a elevação da qualidade do ensino e a construção de um ambiente de aprendizagem mais personalizado, inclusivo e eficiente.

a. Agrupamento por Nível de Habilidade

Agrupar por nível de habilidade consiste em organizar os estudantes de acordo com seu grau de conhecimento e competências em uma área específica, permitindo que o ensino seja ajustado às necessidades de cada um. Esse processo pedagógico difere das práticas tradicionais, que costumam considerar apenas a série ou a idade do aluno, e visa oferecer intervenções mais eficazes e alinhadas ao perfil de aprendizagem de cada grupo.

O agrupamento estratégico de alunos pode atender a diferentes propósitos pedagógicos. Ao formar grupos com habilidades similares, evita-se discrepâncias extremas que dificultam a gestão da sala de aula e comprometem o atendimento aos alunos com atrasos mais severos (ALVES; SOARES, 2007; BETTS; SHKOLNIK, 2000). Por outro lado, grupos heterogêneos podem ser igualmente valiosos, permitindo que alunos com maior facilidade em certas áreas apoiem colegas em dificuldades, promovendo o aprendizado colaborativo. Essa abordagem cria um ambiente de troca de conhecimentos, onde cada grupo recebe desafios adequados e os estudantes se beneficiam de diferentes perspectivas (LECHUGA; DOROUDI, 2023).

A seguir são apresentadas algumas possibilidades de atividades pedagógicas para ensino de matemática utilizando agrupamento por nível de habilidade:

- Enturmação por nível de habilidade: Em uma sala de aula de matemática, isso significa que os alunos não seriam necessariamente agrupados de acordo com sua idade ou série escolar, mas sim com base no seu domínio de conceitos e habilidades matemáticas. Por exemplo, um aluno que tem dificuldades com operações básicas de aritmética pode ser colocado em uma turma focada em reforço desses conceitos, enquanto um aluno que já domina essas operações pode ser inserido em uma turma que trabalhe conteúdos mais avançados, como álgebra ou cálculos geométricos.

Esse tipo de organização permite que o professor adapte sua abordagem de ensino de modo que cada grupo receba a atenção e os recursos necessários para avançar em seu aprendizado, além de possibilitar uma maior personalização do ritmo de ensino. No caso da matemática, onde os conceitos são muitas vezes sequenciais e interdependentes, a agrupamento por nível de habilidade facilita o acompanhamento

do progresso dos alunos e pode resultar em um aprendizado mais eficiente e eficaz para todos.

- Planos de Ensino Diferenciados: As informações coletadas por meio de avaliações podem ser usadas para criar planos de ensino personalizados, direcionados às necessidades específicas de cada grupo de alunos. Por exemplo, estudantes com dificuldades em leitura, matemática ou compreensão textual podem receber atividades de reforço e materiais adaptados que atendam diretamente às suas deficiências. Isso torna o processo de ensino mais eficaz, ao focar os esforços pedagógicos nas áreas que realmente necessitam de atenção (WU et al., 2024).

- Rotinas de Estudo Personalizadas: Com base nos dados e posterior agrupamento é possível desenvolver rotinas de estudo individualizadas, estabelecendo, por exemplo, intervalos de revisão frequente para conteúdos onde os alunos apresentem maior dificuldade. Ferramentas como as aulas da plataforma multimídia da prefeitura (MultiRio) ou miniaulas da Khan Academy, atividades de revisão e jogos educacionais podem ser incorporados para reforçar habilidades específicas, tornando o aprendizado mais dinâmico e adaptado ao ritmo de cada estudante (WU et al., 2024).

- Organização de Oficinas Temáticas: Identificar as áreas de maior dificuldade permite a criação de oficinas temáticas que abordem desafios específicos. Oficinas como “interpretação de problemas matemáticos” ou “leitura e compreensão de textos” podem ser estruturadas para atender necessidades detectadas nos dados (LOUREIRO; RODRIGUES, 2024). Além disso, essas oficinas oferecem oportunidades para parcerias com ONGs e iniciativas voltadas para reforço escolar, alfabetização e preparação para competições acadêmicas, como a OBMEP.

b. Criação de Mapas de Assentos

O mapa de assentos em sala de aula pode ser estruturado com base nos dados de desempenho e nas necessidades dos alunos. Posicionar estrategicamente estudantes com habilidades complementares em duplas ou grupos pequenos pode fomentar a colaboração. Simultaneamente, alunos que requerem mais atenção podem ser colocados em locais onde o professor possa monitorá-los e oferecer suporte mais próximo (BETTS; SHKOLNIK, 2000).

A seguir são apresentadas algumas possibilidades de atividades pedagógicas para ensino de matemática em que o mapa de assentos é recurso estratégico para maximizar a aprendizagem:

Formar duplas com um estudante de alto desempenho e outro de baixo desempenho

Atividade: “Explica pra mim”: A atividade tem o objetivo de estimular a colaboração e a consolidação do conhecimento por meio da explicação, nela o estudante com mais habilidade assume o papel de tutor e explica o conceito para o colega com menor habilidade. O professor fornece um conjunto de problemas com níveis progressivos de dificuldade. O estudante com menor habilidade deve tentar resolver os exercícios, enquanto o outro orienta sem dar a resposta. No final, trocam de papel para reforçar a compreensão.

Formar duplas com dois alunos de alto desempenho. Atividade: “Desafio Matemático”: O objetivo aqui é desenvolver o pensamento crítico e explorar múltiplas abordagens para a resolução de problemas, assim o professor deverá propor problemas complexos ou desafiadores. Os estudantes devem resolver de forma conjunta e justificar cada passo do raciocínio. Podem usar diferentes estratégias para resolver o mesmo problema e depois comparar suas abordagens.

Formar duplas com dois alunos de baixo desempenho. Atividade: “Descobrindo juntos”: O objetivo da atividade é criar um ambiente de aprendizado colaborativo e permitir que os estudantes construam confiança sem se sentirem pressionados. Assim o professor deverá fornecer um conjunto de exercícios com exemplos resolvidos. Os estudantes devem identificar padrões nos exemplos e aplicar o conhecimento para resolver problemas similares. O professor deverá circular pela sala para oferecer suporte sempre que necessário.

Grupos pequenos (misturando habilidades). Atividade: “Missão Matemática”. O objetivo aqui é incentivar a cooperação e promover um aprendizado mais estruturado para todos. Cada grupo recebe um conjunto de problemas e cada aluno fica responsável por explicar e resolver uma parte. Os alunos com maior habilidade assumem papéis de facilitadores, ajudando os colegas sem entregar respostas diretas. O grupo precisa chegar a um consenso antes de registrar a resposta final.

Note-se que estas são sugestões de atividades para alavancar grupos intraclasse formados por critérios baseados na habilidade dos estudantes. Nada

impede que o professor aplique, por exemplo, a ideia de "desafio matemático" a estudantes de baixo desempenho – evidentemente adaptando o desafio proposto.

3 - Aspectos Metodológicos

Nesta seção, apresentamos os requisitos e o processo de desenvolvimento do produto educacional. Inicialmente, detalhamos o caso da Prefeitura do Rio de Janeiro e suas avaliações bimestrais. Em seguida, discutimos a estrutura dos dados desagregados hoje disponíveis, bem como as saídas do sistema (relatórios e visualizações) que compuseram a primeira versão do produto.

3.1 - O caso da Prefeitura do Rio de Janeiro

A Secretaria Municipal de Educação do Rio de Janeiro, com o apoio do Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd/UFJF), realiza desde 2021 provas bimestrais padronizadas e censitárias[2] com o objetivo de diagnosticar o resultado das aprendizagens dos estudantes da Educação Fundamental da cidade apoiando o planejamento de ações educacionais.

Com base nos resultados das avaliações, gestores educacionais e professores da rede podem acessar um *dashboard* que apresenta os resultados e agrupa as questões nas habilidades avaliadas, como exemplificado na Figura 1:

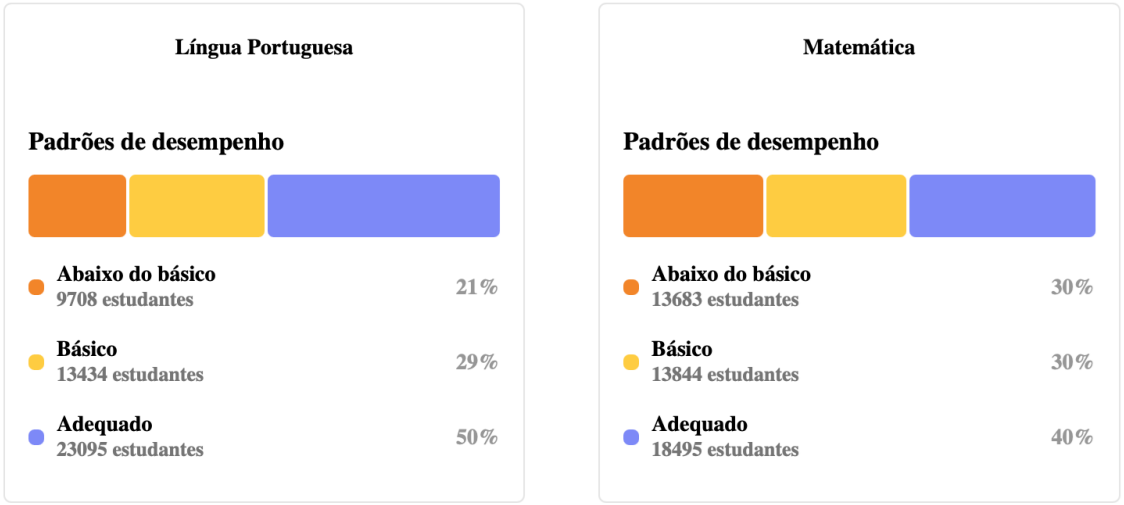


Figura 1: desempenho agregado de todas as escolas da rede municipal do sexto ano no primeiro bimestre de 2024. Fonte: dashboard da prefeitura

[2] As provas são realizadas em todas as cerca de 1.050 Escolas do Ensino Fundamental, com aproximadamente 300 mil alunos do 1º ao 5º ano e em torno de 250 mil do 6º ao 9º ano. Fonte: http://www0.rio.rj.gov.br/sme/numeros.htm

Estas provas produzem dados administrativos que, embora não estejam abertos à sociedade, estão disponíveis hierarquicamente aos gestores de secretaria, escola e professores.

Apesar de ser de alta qualidade, o dashboard oferecido à gestão está focado principalmente em dados agregados, exigindo que todo o acesso granular seja realizado manualmente. Nosso objetivo com este produto educacional é permitir que gestores e professores formem agrupamentos de forma fácil e ágil conforme necessidade do usuário e complementar o dashboard da prefeitura, sem duplicar funções já oferecidas.

3.2 - A estrutura da base de dados

Para atingir seus objetivos a ferramenta acessa os resultados desagregados da avaliação, os quais serão obtidos pelo próprio gestor. Cada avaliação gera diversas tabelas em formato CSV[3], sendo a tabela que contém as notas de cada estudante, HABILIDADES_DESEMPENHO_ESTUDANTE, a mais relevante. Para cada disciplina avaliada (Língua Portuguesa ou Matemática), em cada bimestre (1º ao 4º) e para cada ano escolar (do 1º ao 9º ano), é gerada uma tabela de notas. Assim, em uma escola típica de Ensino Fundamental II, ao final do ano, haverá 40 tabelas de notas, cada uma contendo entre 200 e 300 alunos.

A seguir são descritas as variáveis da tabela referente à prova de Matemática do primeiro bimestre de 2024 aplicada ao 6º ano:

1. Rede – Identifica a rede de ensino à qual a escola pertence. Neste caso, é "MUNICIPAL", por se tratar da Prefeitura do Rio;
2. Etapa – Define a etapa de ensino na qual o estudante está matriculado. Neste caso, seu conteúdo é "Ensino Fundamental de 9 anos – 6º ano";
3. Componente Curricular – Especifica a disciplina da avaliação. No caso, a sigla "MT" indica que a disciplina é Matemática;
4. Código da Turma – Número identificador único da turma à qual o estudante pertence;
5. Estudante – Nome do aluno avaliado. A ferramenta está planejada para ser usada pelos próprios usuários que têm acesso a esta informação;

[3] Sigla do inglês Comma separated values, maneira de escrever dados tabulares em um arquivo de texto sem formatação, separando seus campos por vírgulas (ou outro caractere qualquer – a prefeitura utiliza ";" nestes dados).

6. Avaliado – Indica se o estudante participou da avaliação. “Sim” ou “Não" são os valores possíveis;
7. Padrão de Desempenho – Classificação do desempenho do aluno na avaliação geral. Os possíveis níveis incluem:
    - Abaixo do Básico – Indica que o estudante apresenta grandes dificuldades na disciplina. Até 29% dos acertos possíveis.
    - Básico – Aluno com domínio parcial dos conteúdos avaliados. Entre 30% e 69% dos acertos possíveis.
    - Adequado – Aluno com bom domínio dos conteúdos avaliados. 70% ou mais dos acertos possíveis.
8. Habilidades (H 01 a H 16) – Cada uma das colunas H 01 a H 16 representa uma habilidade matemática avaliada na prova. Os valores apresentados seguem o formato "X / Y", onde:
    - X é o número de acertos do aluno naquela habilidade.
    - Y é o total de questões aplicadas para avaliar aquela habilidade. Exemplo:
      Se um estudante tem "1 / 2" na habilidade H 04, significa que ele acertou 1 questão das 2 aplicadas para essa habilidade.

As descrições das habilidades são informadas na tabela DADOS_ACERTO_POR_HABILIDADE, que especifica textualmente cada habilidade da prova e apresenta o nível médio de proficiência obtido pela turma. Neste caso, as 16 habilidades avaliadas nesta prova de Matemática para o 6º ano do Ensino Fundamental, foram:

1. H 01 – Reconhecer, nomear e comparar figuras geométricas tridimensionais (cubo, bloco retangular, pirâmide, cone, cilindro e esfera), relacionando-as com objetos do mundo físico. Nível: Alto
2. H 02 – Identificar e registrar características do sistema de numeração decimal: base 10 e valor posicional. Nível: Médio Alto
3. H 03 – Ler, escrever, comparar e ordenar números naturais até 99.999. Nível: Alto
4. H 04 – Associar pares ordenados de números naturais a pontos do plano cartesiano no 1º quadrante. Nível: Médio Baixo

5. H 05 – Interpretar representações gráficas (croquis, itinerários, mapas e maquetes), utilizando elementos posicionais: em cima/embaixo, entre, na frente/atrás, ao lado, perto/longe, direita/esquerda etc. Nível: Médio Alto
6. H 06 – Elaborar e resolver situações-problema com números naturais, envolvendo as operações de adição e subtração e suas diferentes ideias: juntar, acrescentar, tirar, completar e comparar. Nível: Médio Baixo
7. H 07 – Identificar características dos quadriláteros e classificá-los em relação aos lados. Nível: Médio Alto
8. H 08 – Identificar características dos triângulos e classificá-los em relação às medidas dos lados. Nível: Baixo
9. H 09 – Calcular potências como produto de fatores iguais (expoente e base naturais). Nível: Médio Baixo
10. H 10 – Realizar adições e subtrações de números fracionários com o mesmo denominador. Nível: Médio Alto
11. H 11 – Identificar os divisores comuns de dois números naturais e reconhecer o máximo divisor comum (MDC). Nível: Baixo
12. H 12 – Identificar os múltiplos comuns de dois ou mais números e reconhecer o mínimo múltiplo comum (MMC). Nível: Baixo
13. H 13 – Reconhecer o significado das frações como ideias de parte-todo (contínuo e discreto), quociente e razão. Nível: Médio Baixo
14. H 14 – Reconhecer, resolver e elaborar situações-problema que envolvam expressões numéricas. Nível: Médio Baixo
15. H 15 – Comparar e ordenar representações fracionárias de uso frequente. Nível: Baixo
16. H 16 – Construir gráficos e tabelas, a partir de situações simples propostas. Nível: Médio Alto

Na prova de Língua Portuguesa foram avaliadas 12 habilidades para o 6º ano do Ensino Fundamental, incluindo o nível de proficiência correspondente:

1. H 01 – Localizar informações explícitas de um texto, literalmente expressas no texto ou dele parafraseadas. (50% - Nível: Médio Baixo)
2. H 02 – Interpretar texto com auxílio de material gráfico diverso (propaganda, quadrinho, foto etc.). (74% - Nível: Médio Alto)
3. H 03 – Estabelecer relações entre partes de um texto, identificando repetições e substituições que contribuem para sua continuidade (substantivos,

pronomes, palavras e/ou expressões mais gerais e mais específicas). (72% - Nível: Médio Alto)

4. H 04 – Identificar a finalidade (função social) de um texto e seu público-alvo. (49% - Nível: Médio Baixo)
5. H 05 – Inferir informações implícitas, seguindo as pistas fornecidas pelo texto como um todo. (48% - Nível: Médio Baixo)
6. H 06 – Inferir o sentido de uma palavra ou expressão no texto percebendo o caráter polissêmico de palavras ou expressões e os diferentes significados que podem assumir em contextos diversos. (86% - Nível: Alto)
7. H 07 – Reconhecer os elementos estruturais da narrativa: situação inicial, complicação (conflito gerador e clímax), desfecho. (45% - Nível: Médio Baixo)
8. H 08 – Reconhecer os elementos da narrativa: Narrador (foco narrativo); personagem (principal e secundários, protagonista e antagonista) e suas características físicas e psicológicas; tempo/espaço da narrativa. (59% - Nível: Médio Baixo)
9. H 09 – Compreender o uso expressivo dos verbos nos textos lidos. (44% - Nível: Médio Baixo)
10. H 10 – Identificar o assunto e o tema (assunto principal) de um texto. (49% - Nível: Médio Baixo)
11. H 11 – Identificar relações lógico-discursivas entre partes de um texto, marcadas por recursos coesivos articuladores de relações de sentido (tempo, lugar, causa, dúvida, comparação, conclusão etc.). (28% - Nível: Baixo)
12. H 12 – Reconhecer os sinais de pontuação e outras notações como marcas de expressividade em um texto, identificando efeitos de sentido de seus usos. (53% - Nível: Médio Baixo)
13. H 13 – Compreender textos não verbais e multimodais, tais como fotos, gráficos, tabelas, tirinhas, propagandas etc. (69% - Nível: Médio Alto)

Além de informações provenientes das avaliações padronizadas, outros dados administrativos podem ser facilmente incorporados, como registros de avaliações qualitativas dos conselhos de classe, a identificação de estudantes beneficiários do Bolsa Família, a participação em programas extracurriculares, entre outros dados coletados rotineiramente pela escola. Esses e outros elementos podem ajudar a frequente tarefa de gerar inteligentemente um grupo de estudantes.

3.3 - Especificação da Funcionalidade Mínima do Produto

Com base na discussão apresentada no referencial teórico, algumas funcionalidades foram priorizadas para compor a primeira versão da ferramenta de gestão inteligente do agrupamento de estudantes. A seguir, abordamos as principais funcionalidades implementadas, considerando tanto seu impacto quanto a viabilidade técnica de execução.

a. Enturmação por Nível de Habilidade

Essa foi a primeira funcionalidade escolhida, pois já havia sido validada pela equipe gestora da escola de origem[4]. A cada bimestre, os resultados das provas de Português e Matemática serão processados somando as notas obtidas por cada aluno em cada habilidade avaliada. A pontuação individual é então calculada como um percentual do total possível, gerando uma média geral da prova e um correspondente nível de habilidade (veja Anexo 7). Em seguida, os alunos são organizados em ordem decrescente de desempenho e redistribuídos em turmas homogêneas. De posse desses dados, equipe gestora poderá reordená-los ou reagrupá-los segundo critérios próprios. Além disso, visualizações adicionais ao dashboard da prefeitura serão criadas para os dados originais.

Em uma versão futura de produção, a ferramenta incorporará ainda outras variáveis relevantes, como o grupo de origem do aluno, turno, afinidades e demais critérios especificados pela escola, garantindo um processo mais flexível e adaptado às necessidades institucionais.

b. Rotinas de Estudo Personalizadas

A partir dos dados de desempenho, será criado um relatório (Anexo 1) que fornece um retorno detalhado ao estudante, indo além da simples atribuição de notas. Com o uso de inteligência artificial, a ferramenta identifica habilidades que ainda precisam ser desenvolvidas, comenta cada uma de acordo com sua dificuldade, e recomenda videoaulas complementares, selecionadas a partir dos acervos da

[4] Aqui é provavelmente importante destacar que, embora a ferramenta tenha sido testada com os dados de uma escola grande, com cerca de 1200 alunos, ela está preparada para funcionar em qualquer escola da rede, bastando ao seu utilizador baixar os dados de sua escola no site da prefeitura e disponibilizá-los ao sistema. Veja o Anexo 8 para os procedimentos necessários.

MultiRio e da Khan Academy, com o objetivo de tornar a revisão de conteúdos mais dinâmica e adaptada ao ritmo de aprendizado de cada aluno.

c. Criação de Mapas de Assentos

Principalmente útil para escolas organizadas em salas ambiente[5], a ferramenta permitirá a criação de mapas de assentos por disciplina, levando em consideração os dados de desempenho dos alunos. Baseando-se em uma abordagem já testada pelo autor em sua prática, os mapas são gerados para dispor os alunos em carteiras duplas, combinando estudantes com maior desempenho com aqueles que necessitam de mais suporte.

Além disso, a ferramenta deve gerar uma visualização imprimível do mapa de assentos, que pode ser fixada na sala para orientação dos estudantes. Em uma versão avançada, a funcionalidade poderia ser aprimorada para considerar necessidades específicas, como a proximidade do quadro para alunos com dificuldades visuais, a alocação estratégica de estudantes que requerem maior supervisão direta ou grupos de tamanhos diferentes.

d. Análises Longitudinais

Essa funcionalidade aborda mais um aspecto não contemplado no dashboard atualmente disponibilizado pela prefeitura. Como requisito mínimo, será implementado um sistema de totalização das notas dos alunos ao longo do ano, permitindo a classificação dos estudantes por nível de desempenho ao final do ano completo.

Adicionalmente, serão criadas visualizações gráficas para ilustrar a evolução das médias das turmas e da escola ao longo do tempo, facilitando o acompanhamento do progresso acadêmico. Essas análises e visualizações visam permitir a gestores e professores a identificação de tendências e o ajuste das estratégias pedagógicas de forma mais informada.

[5] As **salas ambientes** são espaços de aula destinados a um conteúdo ou disciplina específica, equipados com recursos e materiais adequados àquela área de conhecimento. Em vez de cada turma permanecer em uma sala fixa, são os alunos que se deslocam para essas salas de acordo com a matéria que irão estudar.

3.4 - Arquitetura e escolhas de implementação

Apesar de o autor possuir o título de Engenheiro de Sistemas e Computação, grau obtido no século passado, sua experiência pregressa foi de pouca utilidade para o desenvolvimento deste produto educacional. Muito recentemente, a atividade de desenvolvimento de software foi violentamente impactada pela emergência da inteligência artificial generativa. De fato, este evento franqueia a usuários relativamente leigos a faculdade de implementar, ao menos, versões minimamente viáveis dos seus produtos (MVP[6]). De início, adotamos para o MVP uma arquitetura de três camadas simples, como detalhado a seguir.

Camada de apresentação

A camada de apresentação é a interface do usuário ou *front-end*. É a camada de comunicação do aplicativo. Seu principal objetivo é exibir e coletar informações do usuário. Este nível superior é executado em um navegador web, imaginando uma extensão futura para hospedagem na web. A camada de apresentação é desenvolvida usando HTML, CSS e JavaScript, para as atualizações dinâmicas.

Nível de aplicação

A camada do aplicativo ou camada lógica, é o “coração” do aplicativo. Nesse nível, as informações coletadas na camada de apresentação são processadas usando as informações de entrada na camada de dados (sem modificá-las) e gerando novos dados de saída na camada de dados, bem como visualizações na camada de apresentação. Como é desejável que o produto seja multiplataforma, a escolha da linguagem para a camada de aplicação recaiu sobre o Python versão 3.11, uma linguagem interpretada de sintaxe moderna e alta legibilidade, que possui bibliotecas extensas e variadas.

---

[6] Sigla para Minimum Viable Product, ou produto mínimo viável, é uma versão do produto com recursos suficientes para ser utilizável.

Camada de dados

A camada de dados ou *back-end*, é onde as informações processadas pelo aplicativo são armazenadas e gerenciadas. No MVP, em vez de usar um sistema de gerenciamento de banco de dados relacional, como SQLite, MySQL e PostgreSQL – todas opções de software livre – optamos pela alternativa mais simples: arquivos CSV, formato já adotado pela prefeitura. Assim, a saída de dados tabulares é em novos arquivos CSV e as visualizações produzem HTML com JavaScript.

O ambiente de desenvolvimento

De posse dessas especificações e arquitetura, o ambiente de desenvolvimento selecionado foi o Cursor AI, um derivado[7] do Visual Studio Code da Microsoft, mas com extensão para o uso das principais LLMs[8]. Para gerar o código da aplicação dentro do cursor foi adotado o LLM Claude Sonnet 3.5 da Anthropic que nos pareceu funcionar melhor para esta tarefa. A Figura 2, apresenta um instantâneo do ambiente de desenvolvimento.

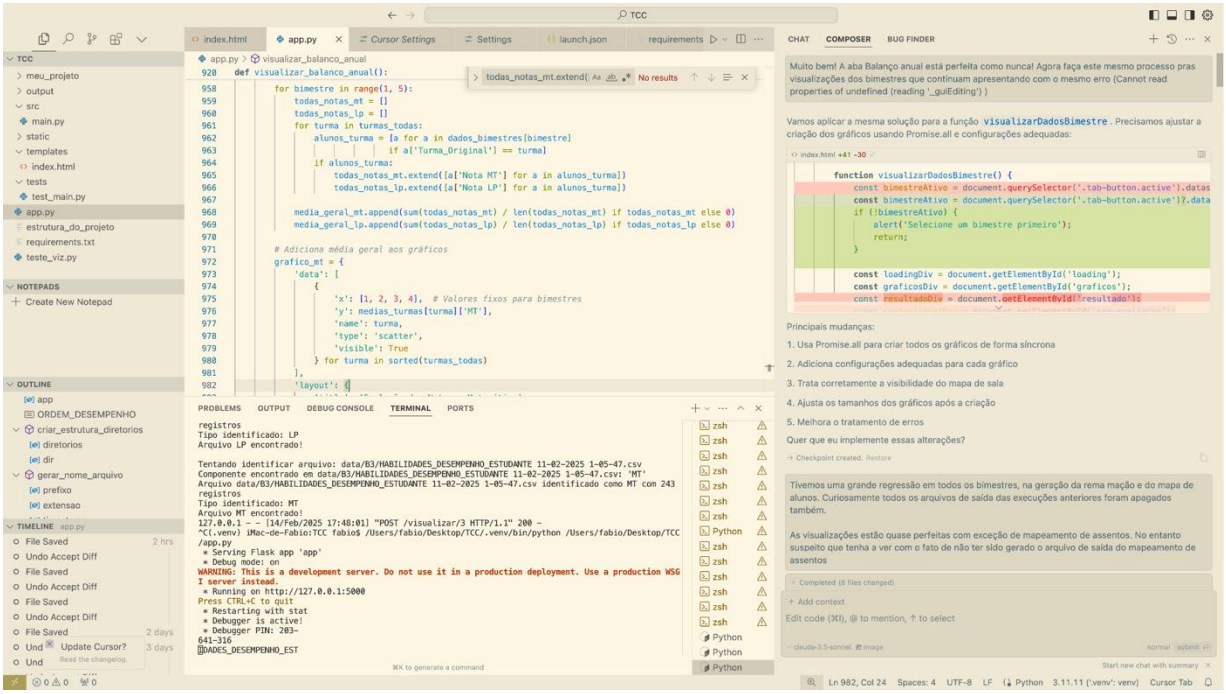


Figura 2 - O ambiente de desenvolvimento Cursor AI

[7] O nome técnico é um “fork”

[8] Sigla de Large Language Model ou grande modelo de linguagem, são redes neurais treinadas com grandes volumes de texto para gerar respostas coerentes e contextuais em linguagem natural.

Em um ambiente deste tipo, a tarefa de desenvolvimento é largamente automatizada: a geração de código é feita a partir de *prompts* em linguagem natural e o principal desafio é a depuração dos erros. Evidentemente, a qualidade dos *prompts* vai definir a precisão da geração do código. Por essa razão, a definição de arquitetura e as opções de ferramentas devem idealmente serem feitas a priori. No caso deste produto educacional, as informações acima foram subsídio para conversas com outra LLM, o ChatGPT-4o. Quando as principais funcionalidades, arquitetura desejada e linguagens de desenvolvimento pareciam já compreendidas, pedi que fosse gerado um prompt para o Cursor AI. A Figura 3 apresenta um exemplo do diálogo inicial utilizado para construção da ferramenta.

Eu tenho um novo projeto. Você deve gerar um MVP em Python 3 usando Flask, conforme abaixo:

Objetivo:
1) Ler todos os arquivos CSV que estiverem na pasta `data/`.
2) Para cada CSV:
- Ler as linhas do arquivo usando o módulo csv da biblioteca padrão (não usar pandas).
- Ordenar os alunos pelo campo "Padrão de Desempenho" (ex.: "Abaixo do Básico" < "Básico" < "Adequado"). Se o valor do campo for desconhecido, ignore o aluno, mas mande uma mensagem com seu nome e “sem avaliação”
- Distribuir os alunos em 6 novas turmas (A, B, C, D, E, F) do maior (A) para menor desempenho (F).
- Gerar um novo arquivo na pasta `data/` com o mesmo nome, acrescido do sufixo `_output.csv`. Cada linha deve conter:
- Nome (campo “Estudante” do CSV)
- Turma Original (campo “Código da Turma”)
- Padrão_de_Desempenho
- Nova Turma (A, B, C, D, E, F)
3) Criar uma aplicação Flask com duas rotas:
- GET `/`: renderiza um template HTML simples (index.html) com um botão “Processar”.
- POST `/processar`: dispara a lógica acima, processando todos os arquivos CSV encontrados e gerando os arquivos de saída. Retorna mensagem de sucesso ou erro.
4) Não usar login ou autenticação, apenas esse MVP.
5) Incluir instruções de como rodar o app (ex.: "python app.py" e então abrir `http://127.0.0.1:5000/`).
6) Explicar rapidamente o que cada parte do código faz, para ficar claro como funcionam as rotas e o processamento CSV.
7) Seguir a seguinte estrutura de pastas:
- `app.py`
- `templates/index.html`
- Pasta `data` onde ficam os CSV de entrada e onde serão gerados arquivos `_output.csv`.

Por favor, gere o código completo (para o arquivo `app.py` e para o arquivo `templates/index.html`) e inclua comentários explicando a lógica. Obrigado!

Figura 3 - O prompt inicial do desenvolvimento do produto

Note-se que este *prompt* contempla apenas uma funcionalidade (a Reenturmação), baseando-se unicamente em 3 campos, de apenas um arquivo de entrada (matemática). Além disso, não foi feita nenhuma consideração de design. Com mais algumas interações já foi possível gerar um primeiro relatório simplificado.

A partir deste ponto, o processo de desenvolvimento é similar ao tradicional: inicia-se o ciclo de testes, correções, ampliação do escopo, novos testes e assim

sucessivamente. Para dar uma ideia deste processo, reunimos alguns *prompts* no Anexo 2. Embora seja possível fazer pequenas correções com a IA, a prática mostrou que as alterações direto no código são muito mais vantajosas. Aqui a escolha do Python se mostrou feliz, por ser uma linguagem interpretada e de sintaxe realmente acessível.

4 - Resultados: O Produto Educacional SIGGA

O Produto Mínimo Viável (MVP) descrito na seção anterior foi plenamente implementado e já se encontra em versão alfa: está funcional, gerando todos os relatórios e visualizações previstos, mas ainda operando de forma centralizada no ambiente de desenvolvimento. Foi batizado como SIGGA, acrônimo para Sistema de Geração de Grupos de Alunos. Para sua implantação nas escolas, será necessária a elaboração de documentação e a definição de uma estratégia de distribuição, que pode envolver tanto a geração de um executável para Windows e Linux quanto a hospedagem em um site público na Internet. Nesse último caso, será necessário um controle de acesso por login e a camada de dados precisará ser atualizada para utilizar um banco de dados relacional, visando compartimentalizar os dados provenientes de diversas escolas. Para mais detalhes técnicos sobre o produto e seu uso, veja o Anexo 8. A seguir, descrevemos as funcionalidades principais do MVP em seu estado atual.

4.1 - Obtenção dos Dados

A base de dados administrativa é carregada em formato CSV, considerando uma escola em um determinado ano letivo e série, contemplando quatro avaliações bimestrais de Língua Portuguesa e quatro de Matemática. Como há arquivos auxiliares, por exemplo o dicionário de habilidades, o total é de 32 arquivos CSV por série (oito por bimestre). Também está prevista a utilização contínua do sistema ao longo do ano, à medida que os novos arquivos forem disponibilizados. Nesse contexto, o sistema está preparado para indicar de forma adequada a ausência de qualquer arquivo de entrada, conforme ilustrado na Figura 4.

Figura 4 - SIGGA: Erro ao selecionar o quarto bimestre prematuramente

Uma vez carregados os arquivos adequados, seleciona-se o bimestre e passa-se à escolha da funcionalidade. Os botões Reenturmação e Mapa de Arquivos produzem arquivos CSV de acordo com suas finalidades. Já o botão de Visualizações mostra visualizações das agregações de dados, não disponíveis no Dashboard da prefeitura como ilustrado na Figura 5.

Na Figura 5 são apresentados 4 gráficos: o histograma das notas obtidas para avaliações de Matemática (esquerdo superior) e Língua Portuguesa (direito superior), o gráfico de dispersão relacionando as notas de Língua Portuguesa e Matemática, as diferentes cores representam cada uma das turmas (esquerdo inferior) e gráfico de barras para visualizar a média bimestral de cada turma, empregando cores distintas para representar as pontuações em Matemática, Língua Portuguesa e a média do bimestre. Note-se que quando o cursor é posicionado sobre algum item, uma mensagem explicativa aparece: no caso, a turma e a média de LP da turma 16.

Figura 5 - SIGGA: visualizações dos dados de um bimestre do 7º ano

4.2 - Enturmação por Nível de Habilidade

Os arquivos gerados para Reenturmação e Mapa de Assentos aparecem, respectivamente, nos Anexos 3 e 4. Curiosamente, nem o dashboard nem os arquivos desagregados da prefeitura permitem a comparação, para um mesmo aluno, de suas notas nas avaliações de Língua Portuguesa e Matemática. Por óbvio, não oferecem também a média geral do bimestre e, muito menos, uma ordenação de acordo com esta nota, como mostrado pelo arquivo de saída, Anexo 3. O algoritmo da reenturmação nesse MVP é bastante simples: os estudantes, em ordem decrescente de notas, são alocados a turmas menos numerosas à medida em que cai sua habilidade demonstrada. O quantitativo de turmas da série é mantido inalterado. A premissa é privilegiar os estudantes com maiores atrasos, oferecendo-lhes turmas menos numerosas e, idealmente, os professores mais experientes, visando trazê-los para níveis de desempenho mais elevados. Na visualização a seguir, disponível no

sistema após a reenturmação, é possível notar enorme discrepância entre turmas de um mesmo ano.

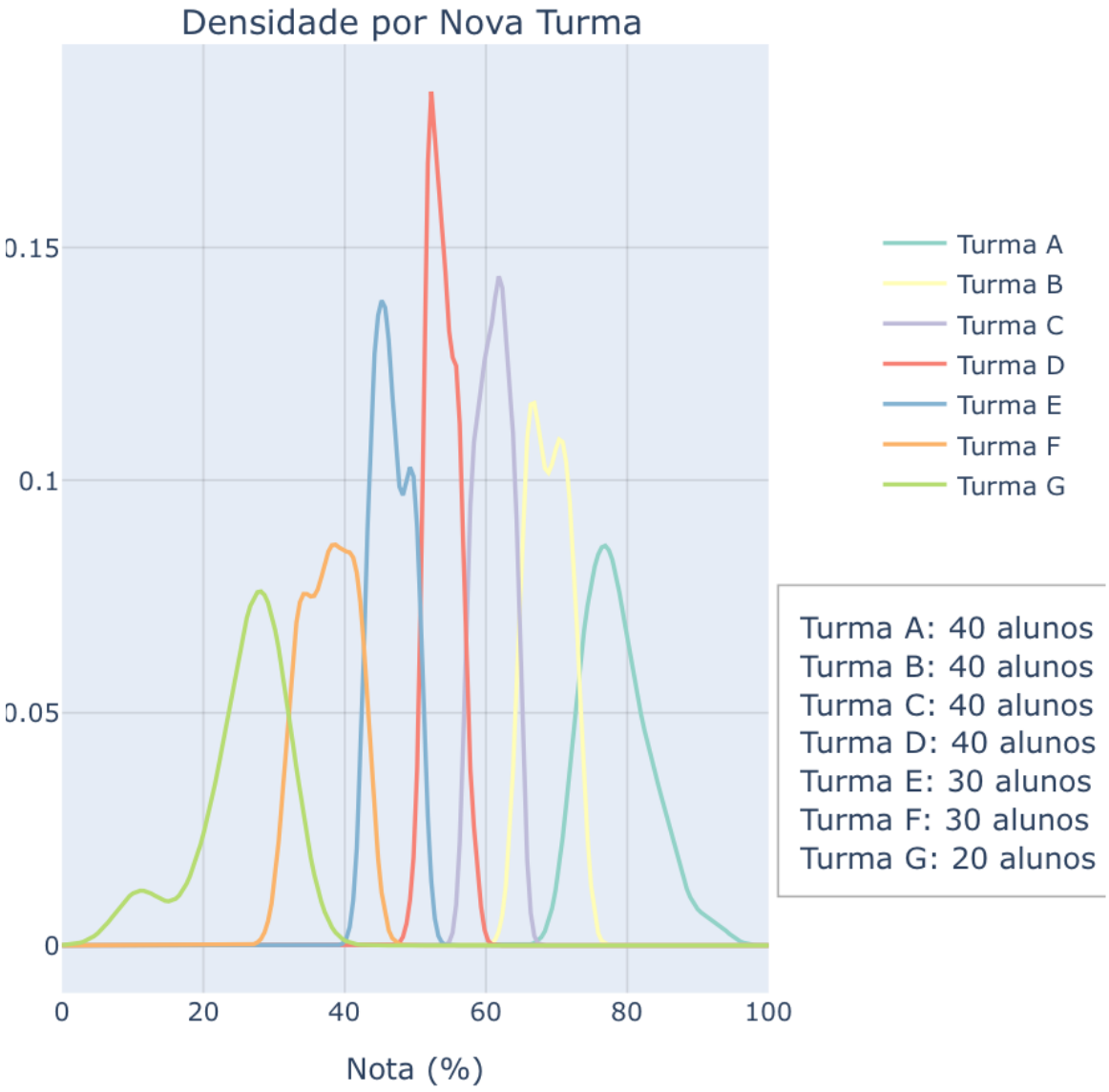


Figura 6 - SIGGA: Densidade de probabilidade das notas das novas turmas e seu quantitativo

4.3 - Criação de Mapas de Assentos

A funcionalidade “mapa de assentos”, como especificado anteriormente, recebeu uma visualização especial. Esta visualização conta com um “semáforo” representando o nível de desempenho do aluno, evidentemente ocultado em uma versão pública, como ilustrado na Figura 6.

Figura 7 - SIGGA: Mapa de assentos imprimível

A média de notas dos estudantes foi utilizada para formar três grupos: verde, com as pontuações mais altas, amarelo, com pontuações intermediárias, e vermelho, com as mais baixas, seguindo a classificação da prefeitura (Anexo 7). A fim de promover a colaboração, os alunos do grupo verde foram emparelhados, sempre que possível, com colegas do grupo vermelho. Essa configuração permite ao professor propor atividades que estimulem a interação, de modo que um estudante auxilie o outro na resolução de problemas.

4.4 - Rotinas de Estudo Personalizadas.

A funcionalidade foi desenvolvida para alunos de alto desempenho, oferecendo um suporte direcionado para pequenas lacunas identificadas em avaliações padronizadas. O foco é no aprimoramento contínuo, garantindo que esses estudantes consolidem seus conhecimentos e elevem ainda mais seu nível acadêmico.

Com base nos resultados da avaliação, a ferramenta identifica habilidades específicas em que o aluno não atingiu pontuação máxima e sugere recursos de estudo complementares, especificamente videoaulas da MultiRio e Khan Academy Brasil. Além disso, reconhece quando a dificuldade pode ter sido causada por um erro de distração, por ser uma habilidade dominada pela maioria, ajustando o feedback para ser mais motivacional e menos instrutivo.

O plano de estudos é gerado automaticamente com chamadas à API[9] do ChatGPT-4o e incentiva o aluno a revisar conteúdos relevantes para seus erros (veja o Anexo 1 para um exemplo). Essa funcionalidade permite que professores ofereçam um acompanhamento mais refinado, transformando dados de avaliação em *insights* acionáveis, que promovem uma aprendizagem mais eficaz e personalizada.

4.5 - Análises Longitudinais

Quando os quatro bimestres estão populados e a base de dados está completa, é possível gerar um "balanço anual” composto de duas funcionalidades: a primeira é um arquivo CSV que consolida os quatro bimestres as notas de português e matemática de e calcula uma média geral, para cada estudante. Este relatório é algo completamente novo para a escola, permitindo rapidamente identificar os alunos de nível elevado de desempenho e aqueles que mais precisam de reforço, no ano letivo completo. Um excerto da listagem está no Anexo 5 e a visualização para estes dados está na Figura 8, consolidada por turma.

[9] API é uma sigla em inglês para Interface de Programação de Aplicações. Ela permite a comunicação e integração de softwares seguindo padrões e protocolos.

Figura 8 - SIGGA: Evolução do desempenho das turmas ao longo do ano

Na Figura 8 são apresentados 4 gráficos: gráfico de linhas considerando os 4 bimestres para cada turma e com a média geral das notas obtidas para avaliações de Matemática (esquerdo superior) e Língua Portuguesa (direito superior), o gráfico de dispersão relacionando as notas de Língua Portuguesa e Matemática, as diferentes cores representam cada uma das turmas (esquerdo inferior) e à direita, um gráfico de barras para ilustrar a nota média anual de cada turma, empregando cores distintas para representar as pontuações em Matemática, Língua Portuguesa e a média geral ao longo do ano. No produto on-line, ao levar-se o mouse sobre qualquer área do

gráfico, aparecem informações adicionais, como por exemplo aqui aluno se refere cada ponto do gráfico de dispersão.

Relatórios nominais de grupos de alunos constituem o principal produto deste sistema. Por isso, as tabelas foram estruturadas de modo a se integrarem facilmente ao fluxo de trabalho das escolas. As listagens de agrupamentos e os relatórios detalhados podem ser exportados em formatos compatíveis com aplicativos amplamente utilizados, como o Excel, o que facilita sua adoção pelos gestores educacionais.

Além disso, o sistema foi desenvolvido para simplificar o trabalho desses gestores, permitindo tanto a personalização de diferentes estratégias de agrupamento quanto a automação de processos que, habitualmente, consomem uma parcela considerável do tempo das equipes escolares.

5 - Discussão

Os resultados apresentados ao longo deste trabalho demonstram como o uso de dados administrativos pode fornecer uma visão mais precisa das necessidades educacionais, tanto no que se refere à identificação de lacunas de aprendizagem quanto na proposição de estratégias concretas para o agrupamento de alunos. A ferramenta desenvolvida (SIGGA) atende aos objetivos definidos no início do estudo, ao indicar a possibilidade de atrasos educacionais sem recorrer a avaliações extras, frequentemente onerosas em termos de tempo e recursos, e o de oferecer uma prova de conceito para intervenções pedagógicas amparadas em informações já disponíveis nas bases governamentais.

A literatura reforça a relevância de estratégias que priorizem a aquisição de competências básicas de leitura e de habilidades matemáticas, destacando a dependência entre essas duas áreas (OLIVEIRA; BORUCHOVITCH; SANTOS, 2008; KORHONEN; LINNANMÄKI; AUNIO, 2012). A evidência de que a proficiência na língua nativa impacta diretamente o desempenho em matemática embasa iniciativas de enturmação e programas de reforço, algo que o sistema SIGGA se propõe a viabilizar de modo prático e eficaz.

A ferramenta desenvolvida também dialoga com autores que enfatizam o potencial das bases de dados institucionais para orientar políticas públicas educacionais (BETTS; SHKOLNIK, 2000; GATTI, 2004). Os resultados alcançados até o momento indicam que a integração desses dados em um produto de fácil uso para professores e gestores escolares favorece a personalização do ensino, a formação de turmas mais adequadas às necessidades dos estudantes e a adoção de intervenções individualizadas, contribuindo para a redução de desigualdades educacionais. Ademais, o recurso às técnicas de inteligência artificial e de mineração de dados educacionais (BAKER et al., 2011; RODRIGUES et al., 2024) abre caminho para o aprimoramento contínuo do sistema, ampliando o impacto de ações de reforço, tutoria e mapeamento de potenciais talentos nas escolas.

Estudos futuros da academia com acesso aos dados de educação da cidade do Rio de Janeiro, com aplicação de técnicas de Mineração de Dados Educacionais (EDM) certamente fornecerão *insights* valiosos sobre os determinantes do desempenho educacional, permitindo uma abordagem mais informada e direcionada na gestão educacional. Esses métodos não apenas identificam alunos que necessitam

de apoio adicional, mas também proporcionarão uma base para o desenvolvimento de políticas educacionais mais eficazes.

5.1 - Sugestões para Estudos Futuros

A partir dos resultados alcançados e das discussões levantadas, algumas linhas de pesquisa são propostas a seguir para aprofundar e expandir o uso de dados administrativos na gestão da aprendizagem:

Exploração do Vínculo entre Literacia e Numeracia

Embora este estudo aponte para a correlação entre leitura e desempenho matemático, seria valioso empreender investigações mais sistemáticas e de maior abrangência que examinem a causalidade. Por exemplo, pesquisas quase-experimentais poderiam verificar se intervenções específicas em letramento contribuem, de fato, para uma melhoria consistente e duradoura na proficiência em matemática, sobretudo no Ensino Fundamental. Um arranjo proposto poderia ser limitar a intervenção a apenas um dos turnos, ficando o outro como grupo de controle.

Prevenção da Evasão Escolar

O uso de dados administrativos não se restringe à identificação de lacunas cognitivas. Registros como notas de Conselhos de Classe (CoCs), ocorrências disciplinares e participação em programas sociais, como o Bolsa Família, podem fornecer indícios importantes de risco de abandono escolar. Estudos futuros poderiam construir modelos preditivos capazes de sinalizar a probabilidade de evasão e embasar intervenções precoces — por exemplo, oferecendo tutoria ou acompanhamento sistemático àqueles alunos identificados como mais vulneráveis.

Integração de Registros Escolares e Indicadores Sociais

A combinação de dados de desempenho em avaliações padronizadas com registros menos estruturados, como ocorrências disciplinares, histórico de repetências, assiduidade, participação em atividades extracurriculares, recebimento de benefícios sociais e resultados de CoCs (Conselhos de Classe), apesar de promissora, apresenta desafios de uniformidade de registro, dificultando sua coleta.

No entanto, essas variáveis podem adicionar dimensões socioemocionais e contextuais à análise, fornecendo uma visão mais completa das trajetórias escolares.

Avaliação de Impacto das Práticas de Enturmação

Embora o sistema proposto de agrupamento por nível de habilidade ofereça benefícios potenciais, faltam estudos controlados que avaliem seu impacto real nos indicadores de aprendizagem, engajamento e continuidade escolar. Experimentos ou estudos de caso conduzidos em diferentes turnos ou excolas poderiam fornecer evidências mais robustas sobre a eficácia (ou limitações) dessas práticas.

Ampliação de Estudos Longitudinais

Um maior período de acompanhamento (por exemplo, acompanhando os mesmos alunos durante todo o ciclo do Ensino Fundamental I e II) permitiria avaliar em profundidade como as estratégias de agrupamento e as intervenções pedagógicas influenciam a progressão escolar e a incidência de fracasso ou evasão. Neste caso, seria fundamental garantir mecanismos adequados de privacidade e segurança na gestão de dados sensíveis.

Ao explorar esses temas, espera-se que o potencial dos dados administrativos seja ainda mais bem compreendido, ampliando a compreensão sobre o vínculo entre literacia e numeracia (vide Anexo 6), bem como aprimorando estratégias de prevenção da evasão escolar em múltiplos contextos educacionais. Dessa forma, os resultados gerados podem subsidiar tanto a melhoria das práticas pedagógicas em sala de aula quanto a formulação de políticas mais eficazes, contribuindo para a promoção de um ensino inclusivo e de qualidade.

5.2 - Limitações do estudo

Esta dissertação oferece um panorama inicial sobre o uso de tecnologias de análise de dados para aprimorar a educação, porém algumas limitações precisam ser destacadas:

Escopo Geográfico e Institucional Restrito

O produto foi concebido com base nos dados administrativos de um contexto específico — a rede municipal do Rio de Janeiro. Outras redes de ensino com diferentes realidades sociais e pedagógicas podem não ter avaliações padronizadas em larga escala, ou serem descontínuas no tempo.

Capacitação e Infraestrutura

A adoção efetiva do sistema pressupõe que gestores e professores possuam algum nível de familiaridade com recursos digitais e tempo para analisar relatórios e implementar intervenções. Em escolas com limitações de infraestrutura (falta de equipamentos, acesso precário à Internet) ou carga de trabalho elevada dos docentes e administradores a utilização da ferramenta pode ser comprometida.

Lacunas na Literatura

A maior parte dos estudos relacionando leitura e desempenho matemático tem sido conduzida fora do Brasil ou em contextos específicos, como crianças com deficiências cognitivas. Isso dificulta a transferência direta dos resultados para o sistema educacional brasileiro, em particular em regiões de alta vulnerabilidade social. Para um primeiro ensaio sobre a relação entre os desempenhos de língua portuguesa e matemática, veja o Anexo 6.

Carência de Longitudinalidade

Ainda faltam pesquisas que acompanhem os mesmos alunos ao longo do tempo para verificar como diferentes intervenções afetam o desempenho em matemática, sobretudo no ensino fundamental. Embora os dados administrativos regionais possam suprir essa lacuna, ainda há barreiras éticas e legais (como considerações de privacidade) que precisam ser superadas.

Produto em Versão Inicial

O SIGGA foi desenvolvido em nível de MVP e encontra-se em versão alfa. Apesar de plenamente funcional em ambiente de desenvolvimento, requer testes mais

extensivos, documentação e estratégias de distribuição para comprovar sua eficácia e facilidade de adoção em cenários reais.

5.3 - Considerações Finais

Embora o presente estudo tenha se limitado a um protótipo (MVP) direcionado a um segmento específico da rede municipal do Rio de Janeiro, seus achados revelam um potencial significativo para aprimorar a gestão da aprendizagem por meio de dados administrativos. O sistema desenvolvido demonstra que, ao centralizar e analisar de forma contínua as informações escolares disponíveis, é possível identificar lacunas de aprendizagem, agrupar alunos de maneira estratégica e personalizar intervenções de acordo com o perfil de cada turma ou estudante — tudo isso sem a necessidade de avaliações e pesquisas adicionais, geralmente onerosas.

Além de agilizar tarefas cotidianas nas escolas, como a definição de turmas de reforço ou a criação de mapas de assentos, a ferramenta contribui para uma cultura de gestão informada por evidências, um requisito fundamental para a elaboração de políticas educacionais mais abrangentes e eficientes. Nesse sentido, a prefeitura detém bases de dados que cobrem centenas de milhares de estudantes ao longo de vários anos, o que abre caminho para análises longitudinais e estudos em larga escala sobre o progresso e o engajamento dos alunos. Explorar tais bases permitiria não apenas dimensionar as necessidades de reforço e promover talentos com maior precisão, mas também avaliar, em profundidade, o impacto de diversas práticas pedagógicas e iniciativas governamentais.

Ainda que o SIGGA se encontre em versão alfa, seus resultados apontam para a factibilidade de ampliar significativamente as práticas de gestão escolar apoiadas em dados. Em última instância, essa abordagem pode embasar políticas públicas capazes de responder mais velozmente às dificuldades dos estudantes e fomentar a equidade e a qualidade do ensino.

Em suma, a coleta e a análise sistemática de dados administrativos, ainda subaproveitados em muitas redes de ensino, podem transformar a forma como escolas e secretarias lidam com o planejamento pedagógico e a gestão de recursos. A pesquisa futura poderá validar e explorar esse potencial, desenvolvendo estudos longitudinais, avaliando intervenções de forma rigorosa e consolidando a gestão informada por dados como uma prática estruturante da educação brasileira.

Referências

ALVES, Maria Teresa Gonzaga; SOARES, José Francisco. **Efeito-escola e estratificação escolar**: o impacto da composição de turmas por nível de habilidade dos alunos. Educação em revista, p. 25-59, 2007.

ALVES, Maria Teresa Gonzaga. **Caracterização das desigualdades educacionais com dados públicos:** desafios para conceituação e operacionalização empírica. Lua Nova: Revista de Cultura e Política, n. 110, p. 189-214, 2020.

BAKER, R.; ISOTANI, S.; CARVALHO, A. **Mineração de dados educacionais**: Oportunidades para o Brasil. Brazilian Journal of Computers in Education, 19(02):03, 2011.

BANERJEE, Abhijit V. et al. **Remedying education**: Evidence from two randomized experiments in India. The quarterly journal of economics, v. 122, n. 3, p. 1235-1264, 2007.

BASSETTO, Camila Fernanda. **Análise regional dos resultados do Saresp**: uma abordagem com modelos hierárquicos. Ciência & Educação (Bauru), v. 27, p. e21063, 2021.

BETTS, Julian R.; SHKOLNIK, Jamie L. **The effects of ability grouping on student achievement and resource allocation in secondary schools**. Economics of Education Review, v. 19, n. 1, p. 1-15, 2000.

DE CRISTO, Fernando; CANAN, Silvia Regina. **Análise de dados sobre docentes e técnicos administrativos em educação dos institutos federais**. Conjecturas, v. 22, n. 17, p. 942-960, 2022.

FIGLIO, David; KARBOWNIK, Krzysztof; SALVANES, Kjell. **The promise of administrative data in education research**. Education Finance and Policy, v. 12, n. 2, p. 129-136, 2017.

FONSECA, Stella Oggioni da; NAMEN, Anderson Amendoeira. **Mineração em bases de dados do Inep**: uma análise exploratória para nortear melhorias no sistema educacional brasileiro. Educação em Revista, v. 32, p. 133-157, 2016.

GATTI, Bernardete A. **Estudos quantitativos em educação**. Educação e pesquisa, v. 30, n. 01, p. 11-30, 2004.

HARTMANN, Nathan et al. **Automatic classification of the complexity of nonfiction texts in portuguese for early school years**. In: Computational Processing of the Portuguese Language: 12th International Conference, PROPOR

2016, Tomar, Portugal, July 13-15, 2016, Proceedings 12. Springer International Publishing, 2016. p. 12-24.

KORHONEN, Johan; LINNANMÄKI, Karin; AUNIO, Pirjo. **Language and mathematical performance:** A comparison of lower secondary school students with different level of mathematical skills. Scandinavian Journal of Educational Research, v. 56, n. 3, p. 333-344, 2012.

LECHUGA, Christopher G.; DOROUDI, Shayan. **Three algorithms for grouping students**: A bridge between personalized tutoring system data and classroom pedagogy. International Journal of Artificial Intelligence in Education, v. 33, n. 4, p. 843-884, 2023

LOUREIRO, Armando; RODRIGUES, Marta de Oliveira. **Student Grouping**: Investigating a Socio-Educational Practice in a Public School in Portugal. Social Sciences, v. 13, n. 3, p. 141, 2024.

OLIVEIRA, Katya Luciane de; BORUCHOVITCH, Evely; SANTOS, Acácia Aparecida Angeli dos. **Leitura e desempenho escolar em português e matemática no ensino fundamental.** Paidéia (Ribeirão Preto), v. 18, p. 531-540, 2008.

RODRIGUES, Henrique Soares et al. **Artificial Intelligence Algorithms to Predict College Students' Dropout:** A Systematic Mapping Study. ICAART (3), p. 344-351, 2024.

RØED, Knut; RAAUM, Oddbjørn. **Administrative registers–Unexplored reservoirs of scientific knowledge?** The economic journal, v. 113, n. 488, p. F258-F281, 2003.

WU, Siyu et al. A **Comprehensive Exploration of Personalized Learning in Smart Education**: From Student Modeling to Personalized Recommendations. arXiv preprint arXiv:2402.01666, 2024.

Anexo 1 - Rotinas de Estudo Personalizadas – Matemática

(Carta dependente dos dados do aluno gerada automaticamente por IA)

📢 [NOME DO ALUNO], Parabéns pelo seu excelente desempenho! Você alcançou 90% dos pontos possíveis na avaliação de Matemática, um resultado impressionante! Isso demonstra seu domínio sobre a maioria dos conceitos avaliados. Houve apenas alguns pequenos ajustes que podem ser feitos para garantir que você continue evoluindo e refinando seu conhecimento.

Abaixo, você encontrará orientações específicas para três habilidades em que não obteve a pontuação máxima. Para uma delas, acreditamos que tenha sido apenas uma distração, enquanto para as outras duas, sugerimos conteúdo da Khan Academy para reforçar seu aprendizado de maneira prática e direta. Vamos lá! 🚀

🔹 Habilidade 02 – Identificar e registrar características do sistema de numeração decimal: base 10 e valor posicional.

📌 Seu desempenho: Muito provavelmente, um erro de distração.

📝 Mensagem para você:

"Sabemos que você domina o sistema de numeração decimal, então é provável que tenha sido apenas um deslize ao marcar o cartão resposta. Isso acontece até com os melhores! Da próxima vez, lembre-se de revisar suas respostas com calma antes de finalizar a prova. Continue assim, sua base matemática está muito bem estruturada!"

✅ Ação recomendada: Apenas atenção redobrada na revisão final da prova.

🔹 Habilidade 11 – Identificar os divisores comuns de dois números naturais e reconhecer o máximo divisor comum (MDC).

📌 Seu desempenho: Algumas dificuldades com MDC.

📝 Mensagem para você:

"O conceito de MDC é muito útil, especialmente quando lidamos com problemas do dia a dia, como divisão de grupos ou simplificação de frações. Como você já tem uma ótima base matemática, um pequeno reforço pode ajudar a consolidar esse conhecimento de forma ainda mais sólida."

🎯 Reforço recomendado:

- 📚 Aula da Khan Academy: Máximo divisor comum (MDC)
- 🏆 Sugestão de prática: Resolver problemas que envolvem divisão de grupos ou organização de eventos, onde encontrar o maior divisor comum pode facilitar a distribuição.

🔹 Habilidade 15 – Comparar e ordenar representações fracionárias de uso frequente.

📌 Seu desempenho: Pequenos ajustes na comparação de frações.

📝 Mensagem para você:

"Frações podem parecer complicadas às vezes, mas quando entendemos os padrões por trás delas, tudo fica mais fácil. O segredo para comparar frações é pensar em equivalência e encontrar denominadores comuns de maneira eficiente. Com um pouco mais de prática, você dominará esse conceito sem dificuldades!"

🎯 Reforço recomendado:

- 📚 Aula da Khan Academy: Comparação de frações
- 🏆 Sugestão de prática: Aplicar a comparação de frações no dia a dia, como ao dividir uma pizza ou calcular proporções em receitas.

📌 Conclusão

Seu desempenho foi incrível, e essas pequenas melhorias vão te ajudar a ir ainda mais longe! O mais importante é manter essa dedicação e curiosidade para aprender sempre mais. Se precisar de ajuda, não hesite em perguntar aos seus professores ou revisar os conteúdos sugeridos.

✨ Continue brilhando! Estamos muito orgulhosos do seu progresso! 🚀🎯

Anexo 2 - Prompts selecionados para a geração da versão MVP do app

PROMPT A - Resultado de uma longa conversa com o ChatGPT-4o, especificando uma estrutura em três camadas (dados lógica e apresentação) e outros requisitos

Eu tenho um novo projeto. Você deve gerar um MVP em Python 3 usando Flask, conforme abaixo:

Objetivo:

1) Ler todos os arquivos CSV que estiverem na pasta `data/`.

2) Para cada CSV:

- Ler as linhas do arquivo usando o módulo csv da biblioteca padrão (não usar pandas).
- Ordenar os alunos pelo campo "Padrão de Desempenho" (ex.: "Abaixo do Básico" < "Básico" < "Adequado"). Se o valor do campo for desconhecido, ignore o aluno, mas mande uma mensagem com seu nome e “sem avaliação”
- Distribuir os alunos em 6 novas turmas (A, B, C, D, E, F) do maior (A) para menor desempenho (F).
- Gerar um novo arquivo na pasta `data/` com o mesmo nome, acrescido do sufixo `_output.csv`. Cada linha deve conter:
  - Nome (campo “Estudante” do CSV)
  - Turma_Original (campo “Código da Turma”)
  - Padrão_de_Desempenho
  - Nova_Turma (A, B, C, D, E, F)

3) Criar uma aplicação Flask com duas rotas:

- GET `/`: renderiza um template HTML simples (index.html) com um botão “Processar”.
- POST `/processar`: dispara a lógica acima, processando todos os arquivos CSV encontrados e gerando os arquivos de saída. Retorna mensagem de sucesso ou erro.

4) Não usar login ou autenticação, apenas esse MVP.

5) Incluir instruções de como rodar o app (ex.: "python app.py" e então abrir `http://127.0.0.1:5000/`).

6) Explicar rapidamente o que cada parte do código faz, para ficar claro como funcionam as rotas e o processamento CSV.

7) Seguir a seguinte estrutura de pastas:

- `app.py`
- `templates/index.html`
- Pasta `data` onde ficam os CSV de entrada e onde serão gerados arquivos `_output.csv`.

Por favor, gere o código completo (para o arquivo `app.py` e para o arquivo `templates/index.html`) e inclua comentários explicando a lógica. Obrigado!

PROMPT B - Para resolver dificuldades com a leitura do arquivo de entrada

Vou te passar o formato do CSV aqui. Ajuste o programa de acordo. Você só precisa obter os campos Código da Turma;Estudante;Padrão de Desempenho

Rede;Etapa;Componente Curricular;Código da Turma;Estudante;Avaliado;Padrão de Desempenho;H 01;H 02;H 03;H 04;H 05;H 06;H 07;H 08;H 09;H 10;H 11;H 12;H 13;H 14;H 15;H 16

MUNICIPAL;ENSINO FUNDAMENTAL DE 9 ANOS - 6º ANO;MT;1656301 ;NOME1 ;Sim;Adequado; 2 / 2; 1 / 1; 1 / 1; 1 / 1; 2 / 2; 2 / 2; 1 / 2; 1 / 1; 3 / 3; 2 / 2; 0 / 1; 0 / 1; 1 / 2; 1 / 1; 2 / 2; 1 / 2

MUNICIPAL;ENSINO FUNDAMENTAL DE 9 ANOS - 6º ANO;MT;1656301 ;NOME2 ;Sim;Abaixo do Básico; 1 / 2; 1 / 1; 1 / 1; 1 / 1; 0 / 2; 2 / 2; 2 / 2; 0 / 1; 2 / 3; 1 / 2; 0 / 1; 0 / 1; 1 / 2; 0 / 1; 1 / 2; 1 / 2

PROMPT C - Depois de várias interações e testes, começamos a estender o sistema

Vamos criar o segundo botão de processamento. Este que já está pronto, vamos chamar de "Reenturmação"; o outro botão vamos chamar de "Mapa de Assentos". Vamos usar os mesmos arquivos de entrada do anterior só que o algoritmo do mapa de assentos cria duplas de alunos, sem mudá-los de turma, montando a dupla com um aluno de alto desempenho com outro de baixo desempenho. Detalhe do arquivo de saída: o sufixo do botão Reenturmação deve ser Reenturmação; E o de Mapa de Assentos deve ser Mapa. No caso do Mapa crie um campo com D1, D2, ... , Dn para cada dupla proposta. Qualquer duvida pergunte antes de gerar

PROMPT D - Calculando as notas

Muito bem. Agora, em vez de se basear no campo Padrão de Desempenho para ordenar os alunos vamos calcular suas notas, somando as habilidades e calculando um percentual do total. Veja a explicação a seguir de como interpretar os campos.

Habilidades (H 01 a H 16) – Cada uma das colunas H 01 a H 16 representa uma habilidade matemática avaliada na prova. Os valores apresentados seguem o formato "X / Y", onde:

X é o número de acertos do aluno naquela habilidade.

Y é o total de questões aplicadas para avaliar aquela habilidade.

Exemplo:

Se um estudante tem "1 / 2" na habilidade H 04, significa que ele acertou 1 questão das 2 aplicadas para essa habilidade.

o percentual (a nota) deve aparecer nos arquivos de saída após Padrão de Desempenho.

Dúvidas?

PROMPT E - Inclusão de novo arquivo de entrada com as provas de português

Parece muito bom. Agora eu gostaria de ter um segundo arquivo de entrada, exatamente igual no formato a este que você já leu mas que representa a prova de língua portuguesa, para estes mesmos estudantes. Você vai ver que o campo componente curricular vai ter a abreviacao LP(língua portuguesa).

Eu quero que você use este arquivo do mesmo modo que fez com o anterior, mas que passasse a calcular , para cada aluno, a média da nota de matematica (vamos Chama-la de "Nota MT") e a nota de língua portuguesa (vamos Chama-la de "Nota LP")

Dúvidas?

PROMPT E - Exemplo de depuração

Uma coisa que esqueci de mencionar é que o número de habilidades varia por prova. Eu recebi uma mensagem "H13" . Imagino que tem a ver com a o arquivo de português ter apenas 12 habilidades. Você pode verificar se foi isso o problema?

Anexo 3 - Formato do Arquivo de Reenturmação

Nome	Turma_Original	Nota MT	Nota LP	Média	Nova_Turma
ALUNO158	1656305	100	88	94	A
ALUNO91	1656307	92	96	94	A
ALUNO144	1656301	88	96	92	A
ALUNO72	1656303	88	92	90	A
ALUNO36	1656305	92	88	90	A
ALUNO98	1656307	85	96	90	A
ALUNO218	1656307	88	88	88	A
ALUNO178	1656301	81	92	86	A
ALUNO98	1656301	77	96	86	A
ALUNO146	1656303	85	88	86	A
ALUNO93	1656306	85	88	86	A
ALUNO100	1656307	85	88	86	A
ALUNO11	1656306	81	85	83	A
ALUNO102	1656301	77	88	82	A
ALUNO173	1656304	73	92	82	A
ALUNO153	1656305	69	96	82	A
ALUNO249	1656306	73	92	82	A
ALUNO173	1656302	81	81	81	A
ALUNO65	1656302	77	85	81	A
ALUNO156	1656301	88	73	80	A
ALUNO216	1656301	65	96	80	A
ALUNO206	1656302	65	96	80	A
ALUNO10	1656302	73	88	80	A
ALUNO90	1656302	69	92	80	A
ALUNO169	1656303	69	92	80	A
ALUNO154	1656301	73	85	79	A
ALUNO199	1656302	73	85	79	A
ALUNO103	1656305	85	73	79	A
ALUNO205	1656306	77	81	79	A
ALUNO119	1656307	81	77	79	A
ALUNO76	1656301	69	85	77	A

Anexo 4 - Formato do Arquivo Mapa de Assentos

Nome	Turma_Original	Nota MT	Nota LP	Média	Dupla
ALUNO36	1656301	92	85	88	D1
ALUNO186	1656301	46	0	23	D1
ALUNO38	1656301	81	81	81	D2
ALUNO200	1656301	38	19	28	D2
ALUNO226	1656301	73	85	79	D3
ALUNO103	1656301	12	62	37	D3
ALUNO104	1656301	69	88	78	D4
ALUNO68	1656301	50	27	38	D4
ALUNO51	1656301	77	77	77	D5
ALUNO53	1656301	19	58	38	D5
ALUNO63	1656301	69	85	77	D6
ALUNO248	1656301	35	42	38	D6
ALUNO238	1656301	65	88	76	D7
ALUNO99	1656301	35	50	42	D7
ALUNO152	1656301	81	65	73	D8
ALUNO110	1656301	46	42	44	D8
ALUNO114	1656301	62	81	72	D9
ALUNO249	1656301	35	54	44	D9
ALUNO14	1656301	77	65	71	D10
ALUNO223	1656301	42	50	46	D10
ALUNO166	1656301	77	62	70	D11
ALUNO244	1656301	58	46	52	D11
ALUNO85	1656301	65	73	69	D12
ALUNO64	1656301	50	54	52	D12
ALUNO27	1656301	73	62	68	D13
ALUNO13	1656301	46	62	54	D13
ALUNO31	1656301	65	69	67	D14
ALUNO230	1656301	46	65	56	D14
ALUNO218	1656301	62	69	66	D15

Anexo 5 - Formato do Arquivo Balanço Anual

nome	turma	media_geral	MT_B1	MT_B2	MT_B3	MT_B4	LP_B1	LP_B2	LP_B3	LP_B4
ALUNO201	1656307	89.75	81	81	88	92	92	92	96	96
ALUNO133	1656305	88.375	96	81	85	100	69	96	92	88
ALUNO137	1656301	87.5	85	85	92	88	77	92	85	96
ALUNO6	1656306	86.875	100	92	92	88	69	81	85	88
ALUNO231	1656302	86.25	88	88	77	65	88	92	96	96
ALUNO104	1656307	84.625	73	77	92	85	81	92	81	96
ALUNO135	1656302	83.25	85	88	77	69	85	85	85	92
ALUNO102	1656303	83.125	81	77	65	85	77	100	92	88
ALUNO76	1656306	82.375	81	58	85	81	81	96	92	85
ALUNO108	1656301	82.25	85	85	81	88	69	96	81	73
ALUNO114	1656303	82.125	81	77	58	88	81	92	88	92
ALUNO234	1656306	81.75	81	85	88	77	65	85	92	81
ALUNO218	1656307	81.25	73	85	96	85	54	81	88	88
ALUNO169	1656302	80.75	81	81	69	73	73	85	96	88
ALUNO194	1656302	80.625	77	92	88	65	77	88	77	81
ALUNO16	1656303	80.625	92	69	77	69	92	73	88	85
ALUNO48	1656301	80.375	85	77	69	81	62	92	85	92
ALUNO240	1656305	80.375	85	85	62	92	62	81	88	88
ALUNO76	1656302	80.0	73	85	81	77	69	85	85	85
ALUNO35	1656302	79.875	69	81	73	73	88	85	85	85
ALUNO60	1656301	79.375	77	73	73	77	81	73	85	96
ALUNO228	1656306	78.5	81	69	85	85	62	77	81	88
ALUNO117	1656303	78.125	69	77	62	65	88	88	88	88
ALUNO31	1656302	77.875	73	92	77	81	73	77	77	73
ALUNO94	1656303	77.75	73	73	65	65	77	88	96	85
ALUNO162	1656301	77.25	77	65	69	77	77	77	88	88
ALUNO158	1656306	77.25	73	73	73	65	73	88	88	85
ALUNO221	1656306	77.125	85	62	81	73	62	77	85	92
ALUNO196	1656303	76.375	85	80	69	69	85	81	77	65
ALUNO10	1656305	75.875	77	73	65	85	69	69	96	73
ALUNO222	1656303	75.5	92	81	65	81	73	77	73	62
ALUNO160	1656301	75.375	77	73	65	69	77	77	88	77
ALUNO247	1656303	75.375	77	77	69	65	85	65	92	73
ALUNO89	1656304	75.0	73	77	65	62	69	85	81	88
ALUNO243	1656302	74.625	81	81	77	81	46	62	88	81
ALUNO30	1656307	74.625	73	81	85	77	62	73	77	69
ALUNO237	1656305	74.5	73	77	69	81	69	77	77	73
ALUNO223	1656304	74.375	85	65	46	73	69	92	73	92
ALUNO245	1656307	74.125	88	85	73	73	62	62	81	69

Anexo 6 - Exploração preliminar do Vínculo entre Literacia e Numeracia

A seguir fazemos uma primeira aproximação, analisando brevemente as provas de língua portuguesa e matemática do primeiro bimestre de 253 estudantes do 6º ano de uma escola da zona sul da cidade. O objetivo é detectar alunos de desempenho deficiente em matemática supostamente causado por uma literacia inadequada à idade, possibilitando a seleção para classes especiais de reforço de alfabetização.

Inicialmente a base teve que ser limpa de dados incompletos, notadamente estudantes que não realizaram uma das provas, resultando nos 253 casos analisados.

As habilidades selecionadas em língua portuguesa foram aquelas em que os estudantes como grupo se saíram melhor, presumivelmente aquelas que seriam mais elementares. Com referência à lista de habilidades de LP, foram selecionadas as habilidades H02, H03, H06 e H13, avaliadas na coorte como nível de acertos alto ou médio-alto. Enquanto as duas primeiras habilidades corresponderam, cada uma, a uma questão no exame, as duas últimas foram resultado da composição de duas questões de prova.

Assim, foi criado um construto de habilidades variando de 0 a 6 (sendo 6 a soma de todas as questões com acerto) e experimentadas algumas ponderações visando uma maior explicação da performance na prova de matemática. Ao final prevaleceu o construto que simplesmente somava os acertos das 6 questões que compõem as habilidades H02, H03, H06 e H13.

Fizemos algumas regressões lineares exploratórias utilizando as notas de cada estudante para português e matemática com o construto de habilidades, em versões padronizadas e não-padronizadas. As relações entre as variáveis são mostradas nas figuras a seguir.

A correlação entre as notas brutas de português e matemática é de 60,3%. Embora exista uma evidente tendência, há também um enorme espalhamento, com vários estudantes bons resultados em apenas uma das provas (Figura 9).

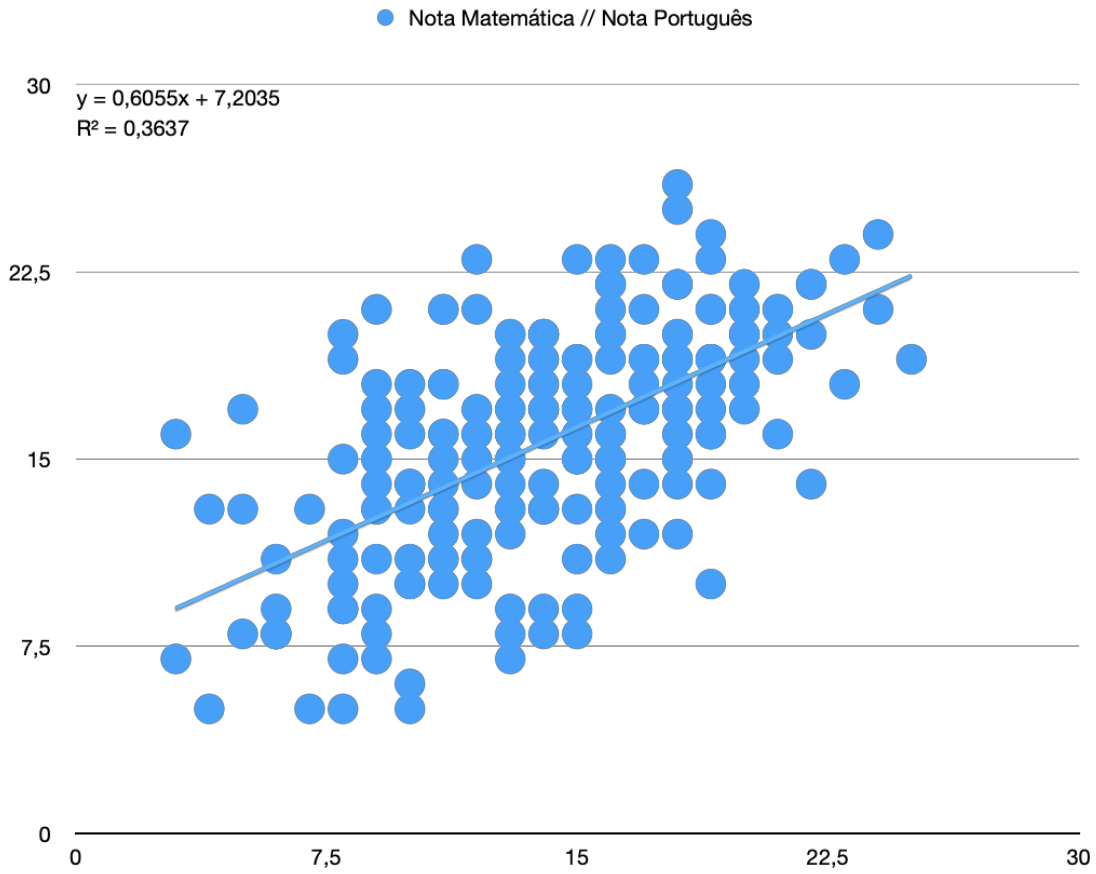


Figura 9 – A dispersão das notas de português e matemática

Quando, em vez das 26 questões de português, passamos a considerar apenas as 6 que constituem as habilidades dominadas pela maior parte da coorte (as questões de maior nível de acerto) era esperável uma queda significativa do poder explicativo do desempenho de matemática. No entanto a correlação passa a ser de 58,7%, permanecendo, portanto, praticamente inalterada (Figura 9).

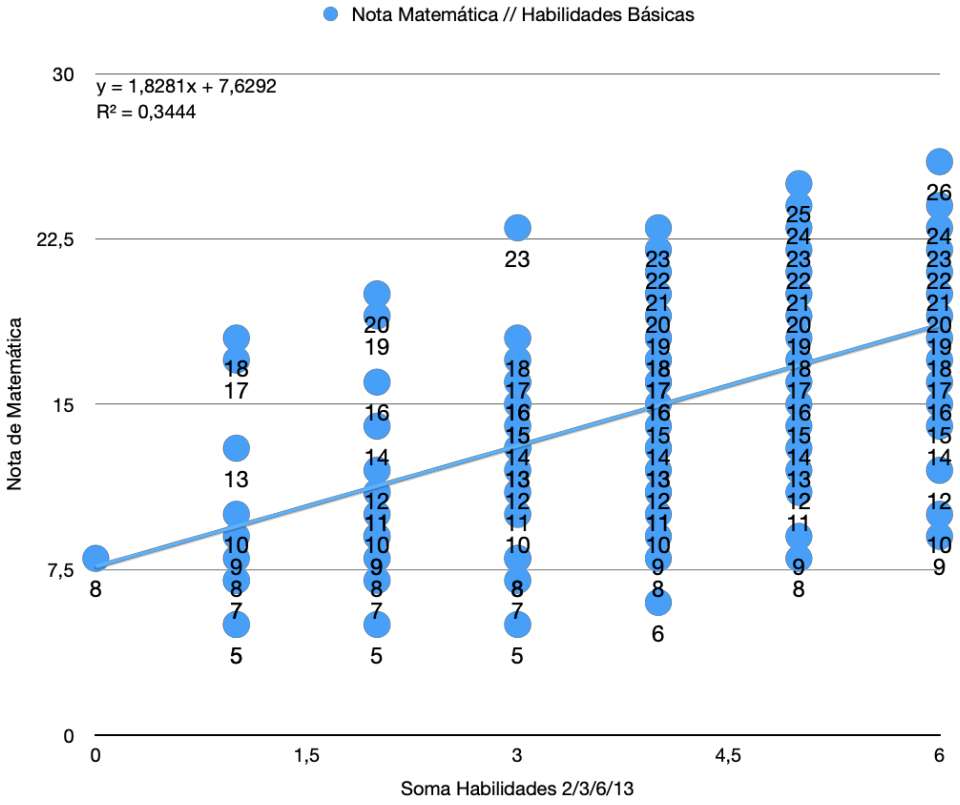


Figura 9 – As notas de matemática pelo construto de habilidades em LP

Embora a correlação seja apenas moderada, ela não cai quando eliminamos 20 das 26 questões da prova de português. Isso sugere que o efeito do conhecimento da língua nativa na matemática está restrito às habilidades mais elementares em língua nativa, como apontado pela literatura.

Anexo 7 - Sistema de Categorização das Notas do SIGGA

Adaptamos o sistema usado pela prefeitura para nos fornecer mais informações. As duas primeiras categorias são obtidas estatisticamente da distribuição binomial, considerando a prova de 26 questões com 4 alternativas de resposta.

1. ⚪ Bola branca (Incoerentes – indícios de erros atípicos)

Critério: Há 80% de confiança de que o estudante não apenas chutou as respostas, mas merece uma investigação adicional.
Limite: Notas abaixo de 20%.

2. ⚫ Bola preta (Chutadores – Acaso, sem conhecimento demonstrado)

Critério: Os acertos do estudante, são compatíveis com a resposta ao acaso, insuficiente para demonstrar um aprendizado real.
Limite: Entre 20% e 31%. (Abaixo de 20%, também pode ter sido chute; acima de 31%, já há evidência de aprendizagem, também com 80% de confiança.)

3. 🔴 Abaixo do Básico (Grandes dificuldades na disciplina)

Critério: O estudante compreende pouco dos conteúdos e apresenta dificuldades significativas.
Limite: Entre 31% e 49%

4. 🟡 Básico (Domínio parcial dos conteúdos)

Critério: O estudante tem um entendimento razoável da disciplina, mas ainda não a domina completamente.
Limite: Entre 50% e 69% (zona intermediária, já indicando um aprendizado estatisticamente sólido, mas ainda incompleto.)

5. 🟢 Adequado (Bom domínio dos conteúdos avaliados)

Critério: O estudante demonstra um conhecimento abrangente.

Limite: Acima de 70% (indica alunos de destaque, que dominam bem a matéria.)

A classificação original da prefeitura é a seguinte:

- Abaixo do Básico – Indica que o estudante apresenta grandes dificuldades na disciplina. Até 29% dos acertos possíveis.
- Básico – Aluno com domínio parcial dos conteúdos avaliados. Entre 30% e 69% dos acertos possíveis.
- Adequado – Aluno com bom domínio dos conteúdos avaliados. 70% ou mais dos acertos possíveis.

Incluímos as duas primeiras categorias para capturar estudantes cujo desempenho é estatisticamente inesperado ou indistinguível de um chute, analisando a distribuição binomial dos acertos. Como cada questão tem quatro alternativas (A, B, C, D), a probabilidade de acerto ao acaso é 25%. Se um estudante chutasse todas as respostas, sua quantidade de acertos seguiria uma distribuição binomial – $X \sim Bin(26,0.25)$ – onde a média esperada de acertos seria 6,5 questões (25% de 26), com um desvio padrão de aproximadamente 2,2 acertos.

A categoria ○ Incoerentes inclui notas abaixo de 20%, que representam a cauda inferior da distribuição binomial. Esse desempenho não apenas indica ausência de conhecimento, mas também sugere que fatores como erros sistemáticos, boicote, desatenção extrema na marcação do cartão ou ansiedade intensa podem estar influenciando as respostas. Estatisticamente, há mais de 80% de confiança de que essas notas não refletem uma resposta meramente aleatória, mas sim um padrão de respostas atípico.

Já a categoria ● Chutadores corresponde ao intervalo entre 20% e 30,77% (5 a 8 acertos em 26), na qual o estudante obteve um desempenho dentro da faixa de variação esperada para respostas ao acaso, ou seja, sem evidência de conhecimento real, pois um estudante chutando poderia atingir essa pontuação — tanto um pouco acima quanto um pouco abaixo do valor esperado de 25% — com uma probabilidade significativa de 63,5%. Adicionalmente, considerando apenas a cauda superior da distribuição, se um estudante obteve até 30,77% de acertos, há 80% de confiança de que ele não mobilizou conhecimento real, apenas respondeu de forma aleatória.

O início da faixa 🔴 Abaixo do Básico é, portanto, a resposta estatística à pergunta "qual é a nota a partir da qual eu tenho certeza de que o estudante sabe alguma coisa e não simplesmente preencheu a prova de forma aleatória?". A partir de 31%, temos 80% de confiança de que o estudante não está apenas chutando, e essa faixa termina imediatamente antes da nota usual de aprovação, 50%. Essa categoria ocupa a primeira metade da classificação "Básico" da prefeitura, mas argumentamos que o termo "Abaixo do Básico" representa melhor esses estudantes, que na maioria das provas ficam abaixo das medidas centrais (média e mediana), indicando dificuldades significativas.

As outras categorias seguem a classificação da prefeitura, reconhecendo o progresso dos estudantes que demonstram aprendizado real a partir do nível 🟡 Básico (50%-69%) e um bom domínio da matéria no nível 🟢 Adequado (70% ou mais).

Figura 10 – SIGGA: Visualizações de ano e turma com as categorias de notas

Anexo 8 - Manual do Usuário SIGGA

1. Visão Geral do Sistema

O SIGGA é um sistema web desenvolvido para auxiliar professores e gestores escolares na análise de desempenho dos alunos, reenturmação e organização de salas de aula. O sistema processa dados de avaliações bimestrais em Matemática e Língua Portuguesa, gerando visualizações, análises estatísticas e recomendações personalizadas.

2. Arquitetura do Sistema

2.1 Tecnologias e Bibliotecas Selecionadas

- Backend: Python/Flask
- Frontend: HTML, JavaScript, CSS
- Visualização de Dados: Plotly.js
- IA para Feedback: OpenAI GPT-4o-mini
- Processamento de Dados: Pandas, NumPy, SciPy

2.2 Estrutura dos Diretórios do Projeto

```
SIGGA/
├── RioEduca/AnoN/
│   ├── B1/        # Dados do 1º bimestre extraídos do RioEduca em Ação
│   ├── B2/        # Dados do 2º bimestre extraídos do RioEduca em Ação
│   ├── B3/        # Dados do 3º bimestre extraídos do RioEduca em Ação
│   └── B4/        # Dados do 4º bimestre extraídos do RioEduca em Ação
│
├── templates/     # Templates HTML
├── static/                # Arquivos estáticos (logomarcas)
├── sigga.py       # Aplicação principal
└── Resultados   # Arquivos de saída
```

3. Funcionalidades Principais

3.1 Processamento de Dados

- Leitura e validação de arquivos CSV
- Cálculo de notas e médias

- Geração de estatísticas por turma
- Processamento de habilidades e notas

3.2 Visualizações

- Distribuição de notas (MT e LP)
- Gráfico comparativo MT x LP
- Médias por turma
- Densidade de notas por turma
- Mapa de assentos
- Balanço anual

3.3 Reenturmação

- Algoritmo de distribuição balanceada
- Visualização da nova distribuição
- Análise comparativa antes/depois

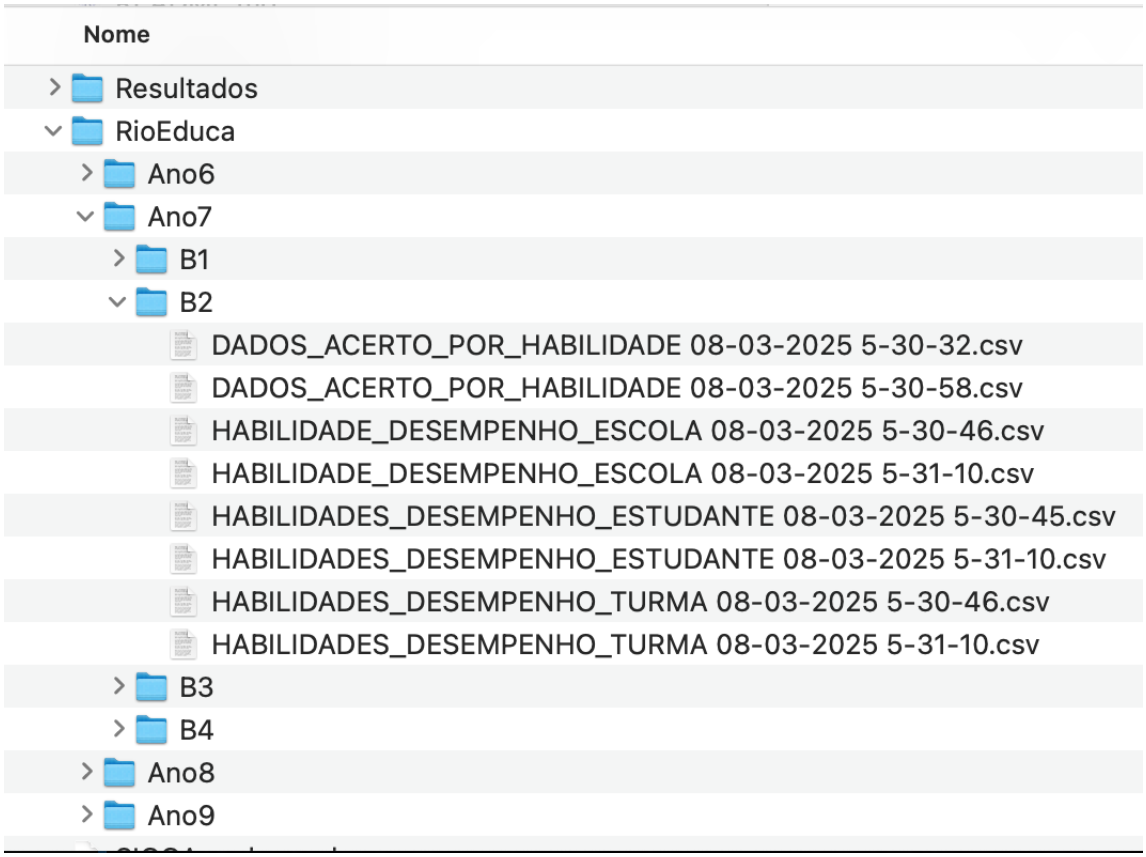
3.4 Autoestudo

- Análise de habilidades a desenvolver
- Recomendações personalizadas
- Carta personalizada com feedback motivador via IA
- Links para recursos externos selcionados pelos professores da escola

4. Guia de Uso

4.1 Primeiros Passos

- Se precisar instalar veja o tópico 7
- Obtenção dos Arquivos: Usando a senha do gestor/professor, obter todos os arquivos CSV para cada ano e matéria (matemática e língua portuguesa) na Plataforma Rioeduca em Ação (https://avaliacaoemonitoramentoriodejaneiro.caeddigital.net)
- Popular as pastas B1, B2, B3 e B4 dentro da pasta AnoX adequada (veja figura a seguir)



4.2 Utilizando o Sistema

Visualização de Dados:

- Selecione o Ano/Bimestre desejado
- Clique em "Visualizar Dados"
- Os gráficos serão exibidos automaticamente

Gerando Mapa de Assentos:

- Selecione o Ano/Bimestre desejado
- Clique em "Gerar Mapa de Assentos"
- Selecione a turma desejada no dropdown

Realizando Reenturmação:

- Selecione o Ano/Bimestre desejado
- Clique em "Realizar Reenturmação"
- Aguarde o processamento
- Visualize os resultados nos gráficos

Autoestudo:

- Selecione o Ano/Bimestre desejado
- Clique em "Gerar Autoestudo"
- Selecione um aluno para ver recomendações
- Clique em "Gerar Feedback" para feedback personalizado

Tratamento de Erros:

- Mensagens de sucesso em verde
- Mensagens de erro em vermelho
- Alertas em amarelo

6. Manutenção e Troubleshooting

6.1 Logs e Monitoramento

Logs são gerados em tempo real na console do terminal
Mensagens informativas para usuários na interface

6.2 Problemas Comuns e Soluções

Arquivos não encontrados:

- Verificar estrutura de pastas
- Confirmar nomenclatura dos arquivos
- Checar permissões de acesso

Erros de Processamento:

- Verificar formato do CSV
- Confirmar presença de todas as colunas necessárias
- Problemas de Visualização:
- Limpar cache do navegador
- Verificar conexão com internet
- Fechar e reabrir página

7. Requisitos do Sistema

7.1 Requisitos de Hardware

- OS: Unix, MacOS ou Windows 8 ou superior
- Memória RAM: 4GB mínimo
- Espaço em disco: 500MB para instalação do Python e pacotes em ambiente virtual

7.2 Requisitos de Software

- Python 3.11 ou superior
- Navegador web moderno
- Conexão com internet para acesso ao Chat GPT via API

7.3 Instalação em ambiente Windows

Para recriar o ambiente virtual Python do zero em uma máquina com Windows, siga a sequência abaixo, executando os comandos no Prompt de Comando (cmd):

A. Navegue até o diretório do seu projeto:

Exemplo:

```
cd C:\Users\fabio\Desktop\SIGGA
```

B. Remova o ambiente virtual antigo (se existir)

```
rmdir /s /q .venv
```

C. Crie novamente o ambiente virtual:

```
python -m venv .venv
```

D. Ative o ambiente virtual:

```
.venv\Scripts\activate.bat
```

D.1. (Opcional) Atualize o pip para a versão mais recente

```
python -m pip install --upgrade pip
```

E. Instale as dependências necessárias:

```
pip install -r requirements.txt
```

F. Execute o aplicativo:

```
python sigga.py
```

G. No navegador, vá até:

http://127.0.0.1:5000

9. Suporte e Contato

Para dúvidas sobre o sistema:

- Contato pelo email: fabio.oliveira@edu.unirio.br

Esta documentação serve como guia para entendimento, utilização e manutenção do SIGGA. Recomenda-se sua leitura atenta antes de contatar o autor.