

PERFIL DOS ALUNOS DO 1º ANO DO IFMG - CAMPUS RIBEIRÃO DAS NEVES VIA FLEXDASHBOARD E ÁRVORE DE DECISÃO PARA PREVISÃO DE DESEMPENHO NA DISCIPLINA DE MATEMÁTICA

Júlia Mota de Souza¹; Gabriela Oliveira²; Marina Muniz de Queiroz³

1 Júlia Mota de Souza, Bolsista CNPq, Curso Técnico em Eletroeletrônica, IFMG Campus Ribeirão das Neves, Ribeirão das Neves - MG; juliassouza.mota@gmail.com

2 Orientadora: Gabriela Oliveira, IFMG Campus Ribeirão das Neves, Ribeirão das Neves - MG; gabriela.oliveira@ifmg.edu.br

3 Coorientadora: Marina Muniz de Queiroz, CEFET Campus Nepomuceno, Nepomuceno - MG; marinamuniz@cefetmg.br

RESUMO

A estatística desempenha um papel fundamental em todas as etapas de uma pesquisa científica, abrangendo desde a identificação do problema até a coleta, tratamento e interpretação dos dados. Além disso, é fundamental para o desenvolvimento do pensamento científico e formação de cidadãos críticos e atuantes. Este trabalho é resultado de um projeto de pesquisa submetido ao Edital 15/2021 da Reitoria do Instituto Federal de Minas Gerais (IFMG) e financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), com o objetivo de proporcionar ao bolsista de iniciação científica um contato mais aprofundado com a pesquisa estatística, desde a preparação do questionário até a coleta de dados e realização das análises por meio do *software* estatístico R, incentivando o interesse pela ciência. O estudo teve como foco a identificação e compreensão do perfil dos alunos do primeiro ano do Ensino Médio no IFMG, *Campus* Ribeirão das Neves, no ano de 2022, com o intuito de aprimorar o planejamento de ações, atividades e projetos no âmbito do desenvolvimento educacional. As questões abordaram temas como idade, gênero, histórico escolar, interesses e motivações, entre outros aspectos. Para apresentação dos resultados de forma a promover a disseminação do conhecimento científico para as comunidades interna e externa do IFMG, utilizou-se o pacote *flexdashboard* do *software* R, que produz uma interface mais dinâmica, atrativa e acessível ao leitor. Algoritmos de árvores de decisão também foram empregados para prever o desempenho dos estudantes na disciplina de Matemática, revelando a relevância do tipo de escola (pública ou privada) frequentada durante o ensino fundamental. Os resultados do projeto podem contribuir para direcionar estratégias educacionais mais eficazes e personalizadas para os discentes do *Campus* e estimular debates e reflexões sobre as práticas educacionais, diversidade e defasagem, visando a melhoria da aprendizagem, o desenvolvimento da cidadania e o sucesso escolar dos alunos.

Palavras-chave: Estatística. Perfil dos alunos. Visualização de dados. Árvore de decisão. Desempenho acadêmico.

INTRODUÇÃO:

A Lei de Diretrizes e Bases da Educação Brasileira (LDB 9394/96) tem como objetivo principal o pleno desenvolvimento do estudante, sua preparação para a cidadania e sua qualificação para o mercado de trabalho (BRASIL, 1996). Nesse contexto, a Estatística desempenha um papel fundamental na Educação Básica, promovendo o desenvolvimento do pensamento científico e a formação de cidadãos críticos e participativos. No âmbito da Matemática no Ensino Médio, a Base Nacional Comum Curricular (BNCC) destaca competências específicas que envolvem a análise de tabelas, gráficos e amostras de pesquisas estatísticas divulgadas em relatórios por diferentes meios de comunicação. Além disso, enfatiza a capacidade de planejar e realizar pesquisas amostrais sobre questões relevantes, utilizando dados coletados de diferentes fontes e comunicando os resultados por meio de relatórios que incluam gráficos e interpretações de medidas de tendência central e dispersão, com ou sem o uso de recursos tecnológicos (BRASIL, 2018). Segundo Cazorla et al. (2017), a capacidade de ler e interpretar informações estatísticas capacita os indivíduos a compreender, avaliar e adotar uma postura crítica diante das notícias frequentemente divulgadas na mídia. Dessa forma, eles se tornam cidadãos capazes de questionar e refletir sobre a veracidade e validade das informações apresentadas.

Por outro lado, a preocupação com as diferenças entre o ensino público e privado tem sido objeto de estudo no Brasil ao longo do tempo. Feijó e França (2021) e Moraes et al. (2022) investigaram o desempenho em Matemática de alunos no final do ensino médio, utilizando os microdados do Exame Nacional do Ensino Médio (ENEM) de 2017, e analisaram os efeitos do contexto familiar e escolar para explicar o melhor

desempenho dos alunos de escolas privadas. A pandemia também é um fator que pode contribuir para as discrepâncias entre alunos de escolas públicas e privadas. Segundo Cavalcanti, Komatsu e Menezes-Filho (2020), as instituições privadas estavam mais preparadas para o acesso a aulas e materiais educacionais durante o período de distanciamento social e parecem ter se adaptado melhor ao ensino à distância em comparação com as instituições públicas, conseguindo fornecer atividades escolares para a maioria de seus alunos. Nesse contexto, Carvalho (2022) conclui que a suspensão das atividades presenciais resultou em um déficit de aprendizado, agravando ainda mais a competitividade entre alunos de escolas públicas e privadas, o que foi evidenciado pela análise das notas na área de Matemática e suas tecnologias no ENEM de 2020. Alves, Queiroz e Oliveira (2022) realizaram uma análise quantitativa utilizando um modelo de regressão linear múltipla para investigar as variáveis relacionadas aos alunos e seus hábitos de estudo de Matemática que aumentam as chances de aprovação nessa disciplina nos primeiros anos dos cursos técnicos no Campus Nepomuceno do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), instituição de ensino semelhante ao IFMG. Os resultados mostraram que os alunos com maior probabilidade de obter um melhor desempenho são aqueles que realizam as atividades de casa, consideram ter tido um ensino fundamental regular ou bom, demonstram facilidade em Matemática e têm pais ou responsáveis com formação de nível médio completo ou superior completo. O modelo não encontrou efeito significativo do tipo de escola frequentada pelos alunos no ensino fundamental.

Além de proporcionar ao bolsista de iniciação científica a oportunidade de se envolver com a análise estatística, este estudo também teve como objetivo compreender e identificar o perfil dos alunos do primeiro ano do ensino médio no IFMG Campus Ribeirão das Neves durante o ano de 2022. Aplicou-se um questionário composto por perguntas separadas em blocos, que abrangiam temas como trajetória escolar, perspectivas para o futuro e uso do tempo em sala de aula. Adicionalmente, empregaram-se algoritmos de aprendizado de máquinas, baseados em árvores de decisão, para realizar previsões sobre o desempenho dos alunos em Matemática, fornecendo informações que podem auxiliar na elaboração de estratégias educacionais mais eficazes. Por meio desses algoritmos, foi possível analisar variáveis relacionadas às características dos alunos e de seus hábitos de estudo e identificar os principais fatores que influenciaram no desempenho acima de 70 pontos dos estudantes nessa disciplina. No contexto desse trabalho, o bom rendimento em Matemática foi estabelecido como a obtenção de nota anual acima de 70 pontos. Isso se justifica pois, em determinados contextos, muitos estudantes são beneficiados no final do ano letivo, seja por concessão de pontos pelo professor, pela recuperação ou por meio do conselho de classe, mesmo quando suas notas estão abaixo dos 60 pontos necessários para aprovação. Assim, utilizando a nota de corte de 70 pontos, foi possível separar os estudantes em dois grupos: os que foram aprovados sem grandes dificuldades ao longo do ano em relação à nota e os que foram reprovados ou aprovados com nota muito próxima ao mínimo necessário.

A análise detalhada do perfil dos estudantes e essas previsões podem orientar professores e gestores na tomada de decisões educacionais embasadas em dados concretos, direcionando esforços para as áreas que necessitam de maior atenção, promovendo um ensino mais personalizado e eficiente, visando atender melhor às necessidades e demandas dos alunos.

METODOLOGIA:

Participaram do estudo 112 alunos que responderam ao questionário aplicado presencialmente no final da terceira etapa em 2022, o que corresponde a 88,4% dos alunos matriculados no primeiro ano do ensino médio dos cursos técnicos integrados de Administração, Eletroeletrônica e Informática. Os dados coletados foram tabulados a partir do Excel e analisados com o *software* estatístico R, com base em medidas descritivas e porcentagens e apresentados por meio de gráficos e tabelas. O *software* livre R é um ambiente computacional integrado e ao mesmo tempo uma linguagem de programação orientada, desenvolvido para análise de dados, realização de cálculos e modelos estatísticos (SILVA et al., 2021).

Para divulgar os resultados obtidos para a comunidade interna e externa ao IFMG, utilizou-se o pacote *flexdashboard* que é uma ferramenta de criação de *dashboards*, que combina recursos estáticos e dinâmicos para apresentar dados de forma visualmente atraente. Ele cria painéis interativos usando o pacote *rmarkdown*, que publica visualizações interativas em JavaScript baseadas em *htmlwidgets* (também funciona com gráficos base, lattice e grid padrão), é flexível e fácil de especificar *layouts*. Os gráficos são redimensionados para preencher o navegador e adaptados para exibição em dispositivos móveis, isso sem a necessidade de possuir habilidades avançadas em programação. Adicionalmente, utilizou-se o pacote *shiny* para conduzir as visualizações de forma dinâmica.

No estudo, empregou-se também algoritmos de *machine learning*, baseados em árvores de decisão, para realizar previsões sobre o desempenho dos alunos. Esses algoritmos, estruturados na forma de árvores, consistem em uma série de decisões lógicas semelhantes a um fluxograma, em que os nós representam a

tomada de decisão com base em um atributo. Os nós se dividem em ramos, e a árvore é finalizada por nós terminais que indicam o resultado resultante de uma combinação de decisões. Para construir a árvore de decisão, é necessário um critério de seleção das variáveis que serão utilizadas como nós, geralmente obtido por meio do cálculo de ganho de informação ou entropia das características da amostra (BRETT, 2020). As características dos alunos e seu comportamento em relação aos estudos, obtidos no questionário, foram utilizados para prever quais alunos teriam um bom desempenho em Matemática. No contexto desse trabalho, o bom rendimento em Matemática foi estabelecido como a obtenção de nota anual acima de 70 pontos. Isso se justifica pois, em determinados contextos, muitos estudantes são beneficiados no final do ano letivo, seja por concessão de pontos pelo professor, pela recuperação ou por meio do conselho de classe, mesmo quando suas notas estão abaixo dos 60 pontos necessários para aprovação. Assim, utilizando a nota de corte de 70 pontos, foi possível separar os estudantes em dois grupos: os que foram aprovados sem grandes dificuldades ao longo do ano em relação à nota e os que foram reprovados ou aprovados com nota muito próxima ao mínimo necessário.

No contexto de aprendizado de máquinas, é crucial avaliar o desempenho do algoritmo em relação ao seu aprendizado. Para isso, os dados são divididos em conjuntos de treinamento e teste (BRETT, 2020). Na modelagem da árvore de decisão, 70% dos dados foram aleatoriamente selecionados para treinar o modelo, enquanto os 30% restantes foram utilizados para validação. O critério de parada estabelecido foi o mínimo de dez observações por folha da árvore.

RESULTADOS E DISCUSSÕES:

Os resultados obtidos indicam que a maioria dos alunos matriculados no 1º ano do ensino médio do IFMG *Campus* Ribeirão das Neves em 2022 é do gênero feminino, representando 56,2% da amostra, enquanto 39,3% se identificaram como do gênero masculino e 4,5% preferiram não responder.

Com relação à cor declarada, a maioria dos alunos do primeiro ano do ensino médio dos cursos técnicos do IFMG *Campus* Ribeirão das Neves em 2022 se autodeclara como preta e parda, somando 69,6% da amostra. Os alunos que se autodeclaram brancos representam 28,6% da amostra, enquanto os que se identificam como amarelos correspondem a 1,8%.

Quanto à faixa etária, as idades mínima e máxima são, respectivamente, 15 e 19 anos. A média de idade dos alunos é de aproximadamente 16 anos.

Os resultados referentes à cidade de residência dos alunos indicaram que 68,8% dos alunos residem em Ribeirão das Neves, enquanto que aproximadamente um terço dos alunos, 31,2%, residem em outras cidades da região. Cerca de metade dos discentes (48,2%) utilizam van, seguidos por transporte próprio (17,8%) e a pé (17,0%). Cruzando os dados da cidade de residência e o meio de transporte utilizado, observou-se que 80,0% dos alunos que residem em Ribeirão das Neves utilizam meios de transporte motorizados, o que pode indicar a necessidade de uma linha de ônibus que contemple o *Campus* como rota.

O estudo também investigou o tempo médio de deslocamento dos alunos até o IFMG em um único trajeto (ida ou volta). Os resultados indicam que 35,7% dos participantes afirmaram gastar menos de 30 minutos, 42,0% gastam entre 30 minutos a uma hora e 22,3% gastam entre uma e duas horas. É importante destacar que cerca de um quinto dos alunos relatou gastar, em média, três horas diárias de deslocamento considerando-se ida e volta.

Sobre a escolaridade do responsável, perguntou-se até que série as mães, madrastas, mães de criação, avós, pais, padrastos e pais de criação estudaram. Ao comparar a escolaridade da mulher responsável pelo aluno com a do homem, constatou-se que o grupo feminino apresentou níveis mais elevados de escolaridade. Além disso, observou-se que a grande maioria dos responsáveis, tanto do gênero feminino quanto do masculino, concluíram pelo menos o ensino médio, com percentuais de 85,7% e 65,1%, respectivamente. O número de alunos que não souberam informar a escolaridade dos responsáveis foi baixo, apenas 1,8%.

Uma das questões investigou a composição familiar. A maioria dos estudantes vive com mãe ou responsável (95,5%) ou com pai ou responsável (76,8%), sendo que a presença feminina é 18,7% maior que a masculina. Verificou-se que 71,4% dos alunos compartilham a residência com irmãos (ãs), enquanto 12,5% convivem com outros parentes.

Quanto aos resultados referentes ao tipo de escola frequentada pelos alunos durante o ensino fundamental observou-se que a maioria dos estudantes, correspondente a 72,3%, cursou todo o ensino fundamental em escolas públicas, enquanto 26,8% frequentaram escolas particulares, seja integralmente ou em parte.

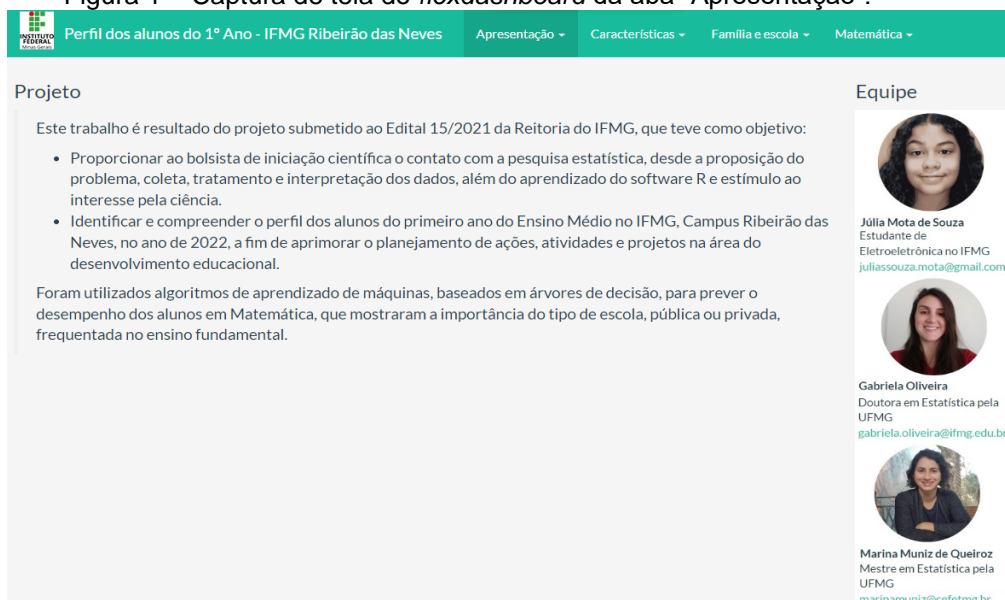
Os alunos foram questionados sobre as razões que os levaram a se matricular no IFMG, sendo apresentadas seis opções possíveis. A opção mais escolhida foi a qualidade e reputação do instituto, correspondendo a 75,0% das respostas. Em seguida, a oferta do curso pretendido foi a escolha de 12,5%

dos alunos. Os demais motivos, como ser o único local onde obteve aprovação, gratuidade, proximidade da residência e outros, foram escolhidos por 11,6% dos estudantes. Apenas um aluno não respondeu a esta questão, o que representa cerca de 0,9% dos participantes da pesquisa.

Nas Figuras 1 a 5 são apresentadas algumas capturas de tela obtidas por meio da utilização do *flexdashboard* como ferramenta de visualização e divulgação deste projeto de pesquisa. A página principal é dividida em quatro abas interativas: “Apresentação”, “Características”, “Família e escola” e “Matemática”. Por sua vez, a aba “Características” possui como subabas “Gênero”, “Cor declarada”, “Idade”, “Cidade” e “Transporte”. “Estudos”, “Escolha IFMG”, “Com quem mora” e “Tempo gasto” são subabas de “Família e escola”. Por fim, tem-se em “Matemática” a subaba “Árvore de decisão”.

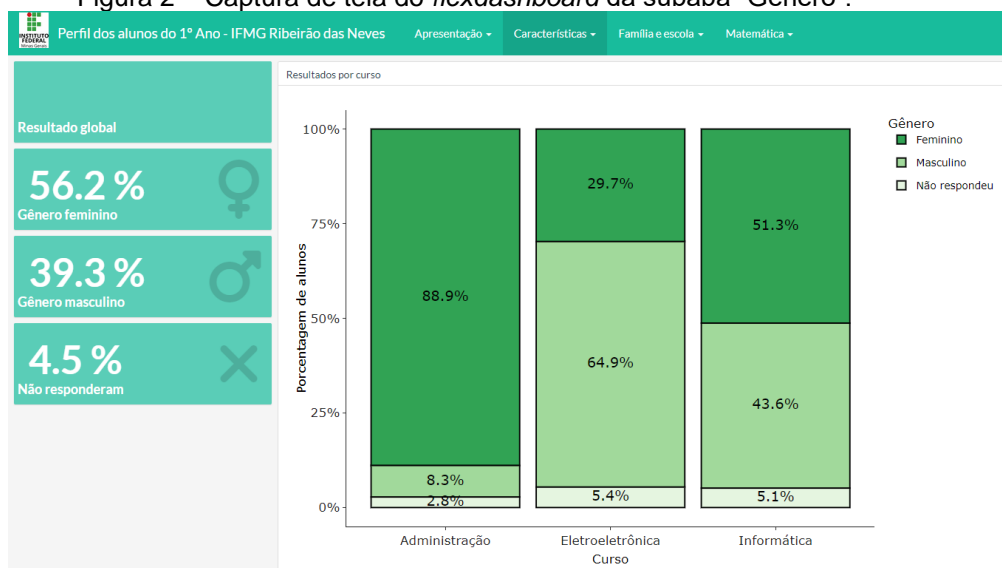
Essa estrutura de abas permite a navegação entre diferentes seções do dashboard para explorar os dados e as análises relacionadas a cada tópico. Destaca-se a importância desse formato de divulgação científica não apenas para a comunidade interna do IFMG, mas também para o público leigo, tornando os resultados mais acessíveis e promovendo a disseminação do conhecimento científico. Os resultados completos podem ser acessados em <https://ifmgneves.shinyapps.io/perfilAlunos1anoIFMG/>. Para uma melhor visualização em *smartphone*, recomenda-se colocar a tela na horizontal.

Figura 1 – Captura de tela do *flexdashboard* da aba “Apresentação”.



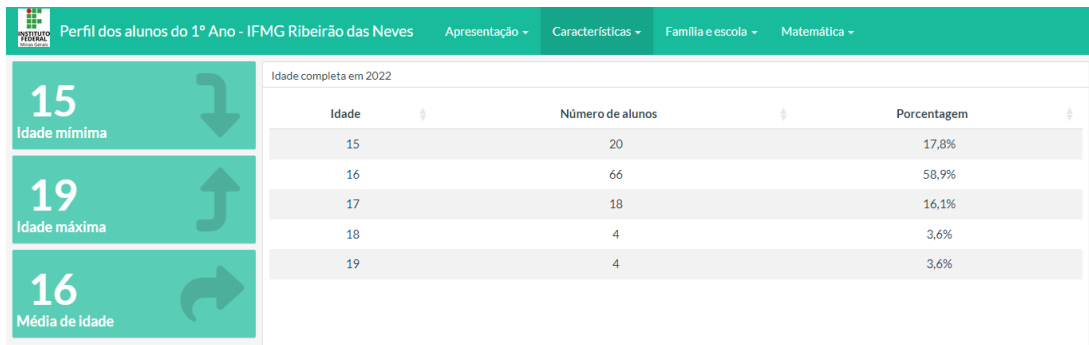
Fonte: Elaborada pelos autores.

Figura 2 – Captura de tela do *flexdashboard* da subaba “Gênero”.



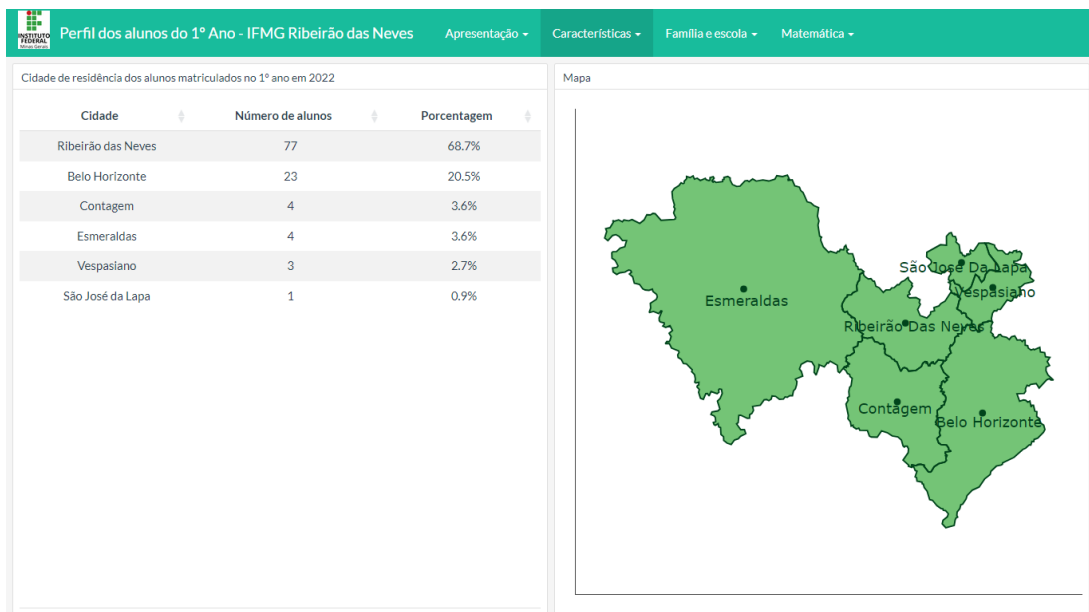
Fonte: Elaborada pelos autores.

Figura 3 – Captura de tela do *flexdashboard* da subaba “idade”.



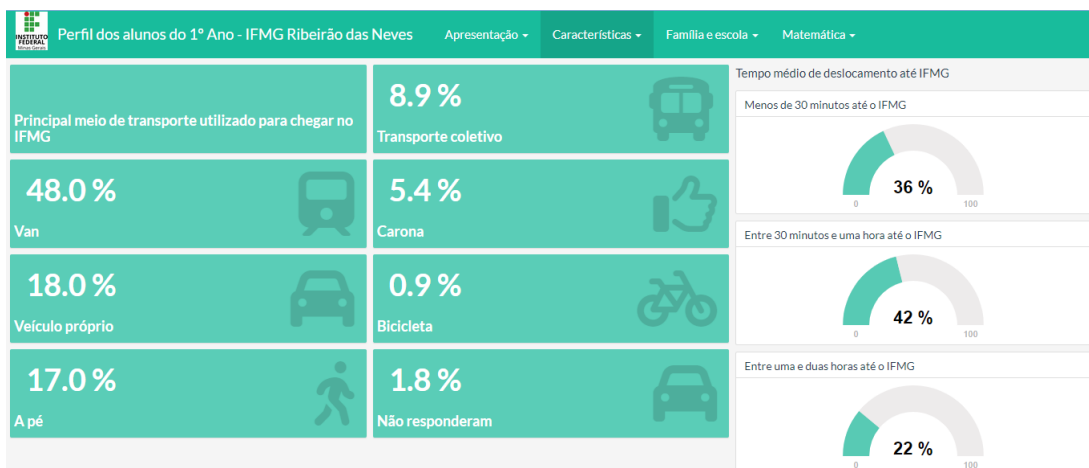
Fonte: Elaborada pelos autores.

Figura 4 – Captura de tela do *flexdashboard* da subaba “Cidade”.



Fonte: Elaborada pelos autores.

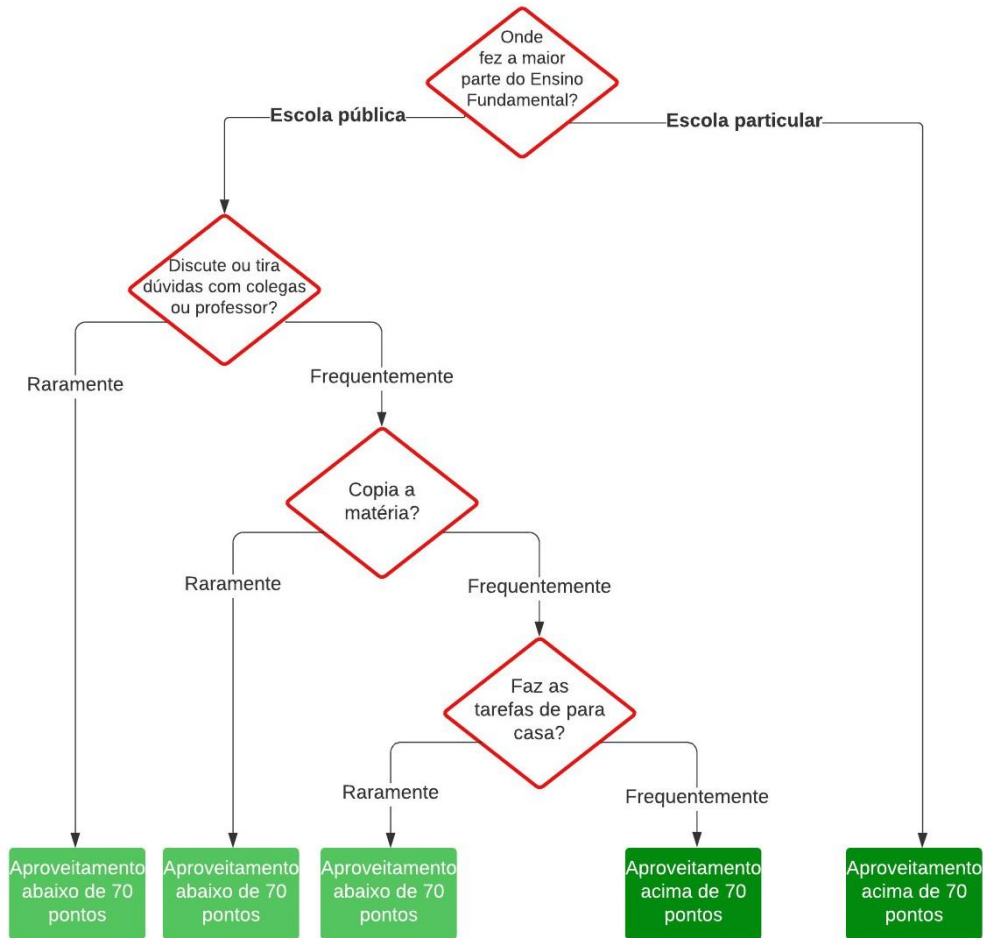
Figura 5 – Captura de tela do *flexdashboard* da subaba “Transporte”.



Fonte: Elaborada pelos autores.

Foi desenvolvida uma árvore de decisão para investigar as características que influenciaram o desempenho dos alunos em Matemática acima de 70 pontos no final do ano, utilizando as variáveis obtidas por meio do questionário. A Figura 6 apresenta o modelo final resultante para a classificação. O primeiro nó da árvore de decisão é a variável "Onde frequentou o ensino fundamental", destacando a relevância do tipo de educação na performance dos alunos. De acordo com o modelo, os alunos que estudaram todo ou a maior parte do ensino fundamental em escolas particulares seriam inicialmente classificados como tendo um bom desempenho, independentemente de outras características e comportamentos relacionados aos estudos.

Figura 6 – Árvore de decisão para classificar os alunos quanto ao aproveitamento em Matemática.



Fonte: Elaborada pelos autores.

O segundo nó ocorre na variável "Discute ou tira dúvidas com outros colegas ou com professor", seguida pelo nó correspondente à variável "Copio no meu caderno a matéria apresentada". Para ambas as variáveis, os alunos que responderam "raramente" não são classificados como tendo um bom aproveitamento.

O último nó é atribuído à variável "Faz as tarefas escolares passadas para casa", e classifica os alunos que frequentemente as fazem como tendo um bom aproveitamento. O modelo foi ajustado utilizando cinco diferentes divisões aleatórias dos dados de teste e treinamento, apresentando uma acurácia média de 82,3%. Além disso, obteve uma sensibilidade média de 83,9% e uma especificidade média de 79,3%, indicando um bom desempenho preditivo.

CONCLUSÕES:

O projeto submetido alcançou o objetivo de proporcionar ao bolsista de iniciação científica o envolvimento com a pesquisa estatística, abrangendo desde a elaboração do questionário até a coleta de

dados, análise e programação no *software R*. Os resultados obtidos foram apresentados de forma interativa e visualmente atrativa, facilitando a compreensão e interpretação dos dados. Esse formato permite que as comunidades internas e externas do IFMG possam compreender e apreciar as descobertas da pesquisa, promovendo a disseminação do conhecimento científico para um público mais amplo, contribuindo para a valorização da ciência e o engajamento da sociedade em temas educacionais.

Além disso, foi estabelecido o perfil dos estudantes do IFMG Ribeirão das Neves e desenvolvido um modelo estatístico baseado em aprendizado de máquinas para prever o bom desempenho dos alunos em Matemática, o que pode subsidiar políticas institucionais, como reforço e aulas de revisão, tanto presenciais quanto online, para alunos que apresentam defasagens no ensino fundamental.

É importante ressaltar que o projeto também incentivou a discente a estudar temas da área de exatas e programação, demonstrando que ações desse tipo podem contribuir para diminuir as desigualdades de gênero nesse campo. Espera-se que este estudo estimule debates e reflexões sobre as práticas educacionais, diversidade e defasagem, visando aprimorar a aprendizagem, promover o desenvolvimento da cidadania e o sucesso escolar dos alunos.

REFERÊNCIAS BIBLIOGRÁFICAS:

ALVES, Thiago Vítor Soares; QUEIROZ, Marina Muniz de; OLIVEIRA, Gabriela. **Análise dos fatores que influenciam o sucesso escolar em matemática dos alunos do ano inicial dos cursos técnicos do Cefet Campus Nepomuceno**. Brazilian Journal of Development, v. 8, n. 1, p. 4870-4879, 2022.

BRASIL, Senado Federal. **Lei de Diretrizes e Bases da Educação Nacional: nº 9394/96**. Brasília: Ministério da Educação, 1996.

BRASIL, Ministério da Educação. **Base Nacional Comum Curricular**. Brasília, 2018.

BRETT, Lantz. **Machine learning with R**. Packt Publishing Limited, 2020.

CARVALHO, Gustavo Almeida. **O impacto da pandemia nas notas do Enem na área: matemática e suas tecnologias**. Monografia (Graduação em Estatística) - Instituto de Ciências Exatas e Biológicas, Universidade Federal de Ouro Preto, Ouro Preto, 2022.

CAVALCANTE, Vitor; KOMATSU, Bruno Kawaoka; MENEZES-FILHO, Naercio. **Desigualdades Educacionais durante a Pandemia**. Policy Paper, n. 51, 2020.

CAZORLA, Irene, et al. **Estatística para os anos iniciais do ensino fundamental**. Sociedade Brasileira de Educação Matemática, Biblioteca do Educador-Coleção SBEM 9, 2017.

FEIJÓ, Janaína Rodrigues; FRANÇA João Mário Santos de. **Diferencial de desempenho entre jovens das escolas públicas e privadas**. Estudos Econômicos (São Paulo), 51(2), 373–408, 2021.

MORAES, Caroline Ponce de; PERES, Rodrigo Tosta; BARBOSA, Tereza Serrano; PEDREIRA, Carlos Eduardo. **Efeito escola a partir de indicadores educacionais: análise entre escolas públicas e privadas no ENEM**. Revista Meta: Avaliação, 2022.

R Development Core Team. **R Markdown Format for Flexible Dashboard**. URL: <https://pkgs.rstudio.com/flexdashboard/>. Acesso em: 2023-05-02.

SILVA, Ana Hermínia Andrade e et al. **Introdução à estatística no software R**. João Pessoa: Editora UFPB, 2021.