

SAEN: UM SISTEMA DE ANÁLISE DE DADOS BASEADO EM RESULTADOS DO ENADE

Breno Brito Cruz*
Kilbert Amorim Maciel**

RESUMO

A análise de dados continua sendo um desafio complexo devido à diversidade de formatos e à imensa quantidade de informações disponíveis. Isso dificulta o processamento dos dados e a obtenção de *insights* relevantes. No caso específico do exame ENADE, que fornece dados históricos em arquivos, é necessário um manuseio manual e estatístico para disseminar as informações, o que prejudica a eficácia das análises realizadas. Este artigo propõe a criação de uma ferramenta *web* que visa aprimorar a experiência na análise dos dados do exame ENADE. Através dessa ferramenta, os usuários poderão visualizar os dados utilizando gráficos de comparação, com foco principal na comparação dos Institutos de Ensino Superior. A ferramenta permitirá consultar os dados das edições de 2014 a 2021 de todas as áreas de avaliação e Institutos de Ensino Superior do Brasil.

Essa abordagem inovadora proporcionará uma forma mais eficiente e intuitiva de explorar e compreender os dados do ENADE. Ao utilizar gráficos de comparação, os usuários poderão identificar padrões, tendências e diferenças entre os Institutos de Ensino Superior, facilitando a análise comparativa dos resultados. Com acesso fácil aos dados de várias edições do ENADE e de diferentes áreas de avaliação, essa ferramenta oferecerá uma visão abrangente do desempenho acadêmico no país, contribuindo para a tomada de decisões informadas e para a melhoria da qualidade do ensino superior.

ABSTRACT

Data analysis remains a complex challenge due to the diversity of formats and the immense amount of available information. This hinders data processing and the extraction of relevant insights. In the specific case of the ENADE exam, which provides historical data in files, manual and statistical handling is required to disseminate the information, thus impairing the effectiveness of the analyses conducted.

* Graduando em Bacharelado em Ciência da Computação, Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), Aracati, Ceará, Brasil. E-mail: brenobritobc@gmail.com

** Mestre em Computação, Docente do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), Aracati, Ceará, Brasil. E-mail: kilbert.maciел@ifce.edu.br

This paper proposes the creation of a web tool aimed at enhancing the experience in analyzing ENADE exam data. Through this tool, users will be able to visualize the data using comparison graphs, with a primary focus on comparing Higher Education Institutions. The tool will allow for accessing data from the 2014 to 2021 editions across all evaluation areas and Higher Education Institutions in Brazil.

This innovative approach will provide a more efficient and intuitive way to explore and comprehend ENADE data. By utilizing comparison graphs, users can identify patterns, trends, and differences among Higher Education Institutions, facilitating comparative analysis of the results. With easy access to data from various ENADE editions and different evaluation areas, this tool will offer a comprehensive view of academic performance in the country, contributing to informed decision-making and the improvement of higher education quality.

1 INTRODUÇÃO

De acordo com (EDUCAÇÃO, 2023), o Instituto Nacional de Estudo e Pesquisas - INEP realiza a aplicação do Exame Nacional de Desempenho dos Estudantes desde 2004, com o objetivo de avaliar o desempenho dos estudantes brasileiros em instituições de ensino superior (IES). O exame é aplicado anualmente e aborda um grupo de áreas de ensino com um intervalo de três anos, sendo uma avaliação obrigatória para os alunos inscritos. Ele é composto pela prova, que irá aferir o desempenho dos alunos de acordo com os conteúdos de um curso de graduação, juntamente com o questionário do estudante, que busca obter informações sobre o contexto formativo do aluno.

Atualmente os dados do exame são disponibilizados no site do INEP, em edições individuais nos formatos XLS(em inglês, Excel Spreadsheet, Planilha do Excel), XLSX(em inglês, Excel Open XML Spreadsheet, Planilha Excel Open XML), CSV (em inglês, Comma-Separated Values, Valores Separados por Vírgulas) e relatórios no formato PDF (em inglês, Portable Document Format, Formato de Documento Portátil) . Nota-se que ao realizar análises nos arquivos Excel ou CSV, é necessário conhecimento analítico e técnico da ferramenta para a busca manual dos dados. Já nos relatórios em PDF, há uma limitação nas informações, já que as mesmas foram escolhidas pelo próprio INEP. Assim, é evidente a necessidade de uma melhoria na forma que os dados são disponibilizados para o público.

A estatística é ferramenta fundamental para que se possa traçar planos sociais e econômicos e projetar metas para o futuro. Técnicas estatísticas avançadas permitem estimar com um bom grau de precisão variáveis como tamanho da população, taxa de emprego e desemprego, índices de inflação, evasão escolar, demanda por determinados bens e serviços, assim como formular planos para atingir as metas programadas de avanço no bem-estar social. Percebe-se que a manipulação de pequenos a grandes volumes de dados é tão usada desde uma corporação, até a área da saúde e educação, e cada vez mais pode ser utilizada em prol da análise e uma possível melhora em algo na sociedade. (IGNÁCIO, 2010)

Este trabalho propõe a criação do sistema SAEN, uma ferramenta *web* que tem como objetivo permitir uma comparação entre o desempenho de duas universidades utilizando técnicas de visualização de dados, permitindo uma melhor análise e tomada de decisão, tanto para a gestão de ensino em relação a preparação dos alunos de ensino superior, quanto para o público externo que busca verificar o nível de qualidade de uma determinada instituição de ensino superior em relação a outra.

A funcionalidade de comparação de universidades permite consultar dados disponibilizados pelo INEP de 2014 a 2021, para em seguida serem criados gráficos comparativos entre as notas e anos, e outros gráficos com dados complementares como quantidades de alunos concluintes e alunos participantes. A consulta tem como base a área de avaliação, as instituições selecionadas, os municípios e por fim a edição inicial e a edição final.

O trabalho está dividido nos seguintes tópicos: A seção 2 consiste na Fundamentação Teórica, a qual explicará os termos utilizados na construção deste trabalho. A seção 3 apresenta trabalhos relacionados que abordam a criação de ferramentas para análise de dados do ENADE. Na seção 4, será detalhada a metodologia da criação da aplicação, incluindo o fluxograma, a coleta de dados, a criação da interface, entre outros aspectos. A seção 5 traz os resultados do trabalho e apresenta o funcionamento do SAEN, além da avaliação da ferramenta. Por fim, na seção 6, temos a conclusão, que abordará o que o trabalho realizou e irá detalhar as funcionalidades para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Conceito Enade

"Os resultados do Enade serão utilizados para fins de cálculo dos Indicadores de Qualidade da Educação Superior"(Diário Oficial da União, 2023), permitindo identificar o nível dos institutos de ensino superior em uma determinada área de ensino.

Com isso, a nota do exame Enade é "a média ponderada das notas padronizadas do respectivo curso de graduação em FG e CE, sendo 25% o peso da Formação Geral e 75% o peso do Componente Específico da nota final."(INEP, 2022)

A equação utilizada utilizada para a composição da nota é apresentada na Equação 1:

$$NC_c = 0.25 \cdot NP_{FG_c} + 0.75 \cdot NP_{CE_c} \quad (1)$$

Fonte: (INEP, 2022)

Onde " NC_c é a nota dos concluintes no Enade do curso de graduação c ; NP_{FG_c} é a nota padronizada em FG do curso de graduação c ; e NP_{CE_c} é a nota padronizada em CE do curso de graduação c ". (INEP, 2022)

Com isso, é gerado o Conceito Enade, que "é uma variável discreta que assume valores de 1 a 5, resultante da conversão da Nota dos Concluintes no Enade do curso de graduação."(INEP, 2022)

Conceito Enade (Faixa)	NCc (Valor Contínuo)
1	$0 \leq NCc < 0,945$
2	$0,945 \leq NCc < 1,945$
3	$1,945 \leq NCc < 2,945$
4	$2,945 \leq NCc < 3,945$
5	$3,945 \leq NCc \leq 5$

Quadro 1. Parâmetros de conversão NCc em Conceito Enade, Fonte: (INEP, 2022)

Conforme o Quadro 1, temos o Conceito Enade Contínuo (NC_c) no qual é uma nota de 0 a 5 sendo um valor decimal, e devido a isto permite uma análise mais específica do resultado. Já o Conceito Enade Faixa permite uma análise de forma mais simples, entre valores de 1 a 5.

O Conceito utilizado para representar as notas nos gráficos da aplicação foi o Conceito Enade Faixa, para uma possível padronização em meio a diferentes tipos de notas.

2.2 Análise de Dados e Tomada de Decisão

A análise de dados é o processo de formação de sentido além dos dados, e esta formação se dá consolidando, limitando e interpretando o que as pessoas disseram e o que o pesquisador viu e leu, isto é, o processo de formação de significado. A análise dos dados é um processo complexo que envolve retrocessos entre dados pouco concretos e conceitos abstratos, entre raciocínio indutivo e dedutivo, entre descrição e interpretação. (TEIXEIRA, 2003).

O crescimento acelerado no volume de informações geradas e uma necessidade do mercado em tomadas de decisão mais rápidas e precisas criaram a necessidade de se estudar formas mais eficientes de gerar e demonstrar conhecimento (MATSUBA; MATTEDI, 2021).

No Brasil, o desafio de analisar e compreender o comportamento dos alunos é muito grande devido a diversidade da população.(BAKER; ISOTANI; CARVALHO, 2011), porém nota-se um potencial no valor dos dados educacionais, permitindo realizar análises dos dados e a partir da análise realizar uma tomada de decisão que pode acarretar em uma melhoria nas notas dos alunos nas avaliações.

Nesse contexto, O SAEN busca permitir uma maneira acessível de realizar análise de dados do exame ENADE que são disponibilizados pelo INEP.

2.3 Visualização de Dados

Com a enorme quantidade de dados que são gerados diariamente, os dados abertos se tornaram cada vez mais comuns em nossas vidas, sendo assim um desafio democratizar as visualizações e realizar análises sobre tais informações."(BORCHARDT et al., 2022)

Com isto surgiu o termo visualização de dados, o qual "possui como objetivo representar visualmente informações quantitativas a respeito dos conteúdos trabalhados, de forma que facilite a compreensão humana, como também, auxilie a descoberta do conhecimento, tendências e

padrões nos dados."(FERREIRA et al., 2022), porém "A eficácia de uma visualização depende do contexto em que é aplicada, definida por usuários, tarefas e dados."(KUNZAYILA, 2021)

"A visualização da informação estuda maneiras de melhorar a interação visual do usuário com a informação para facilitar o entendimento das informações para a tomada de decisão. As visualizações permitem uma retenção maior de conhecimento, um entendimento mais apurado e comparações mais eficazes em relação aos dados brutos. Diferentes formatos visuais como gráficos, infográficos e vídeos podem ser explorados. Porém, é necessário tomar cuidado com a escolha, pois ao invés de facilitar o entendimento, a mesma pode distorcer a informação ou até mesmo dificultar seu entendimento."(TRASSI, 2016). Com isto percebe também que "tornar a apresentação dos resultados visual, fácil de consumir, permitindo consultas, interação e exploração torna a tomada de decisão mais rápida e eficaz"(COSTA, 2017)

Segundo (BOSCARIOLI; VITERBO; TEIXEIRA, 2014) "Há diferentes técnicas de visualização de dados, porém, a grande maioria é limitada pela dimensionalidade da base de dados a ser explorada. Com o passar do tempo, várias técnicas foram desenvolvidas para diferentes tipos de dados e também para diferentes dimensionalidades."

Os dados do SAEN foram disponibilizados em gráficos que mostram as respectivas notas das instituições selecionadas entre intervalo de edições .

2.4 Experiência do Usuário

A experiência do usuário(User Experience), é a "percepção subjetiva feita pelo usuário de um produto. A usabilidade, ao contrário, refere-se ao comportamento de uso e ao propósito do uso. No entanto, para criar fidelidade à marca a longo prazo e manter os usuários satisfeitos, a usabilidade simples já não é mais suficiente. O foco da UX capta fatores emocionais e visa uma experiência positiva e de construção de confiança"(International Organization for Standardization, 2018)

Design de interface preocupa-se com fatores relacionados à maximização da usabilidade e experiência do usuário. Fatores como *web design*, *design* de conteúdo, navegação, acessibilidade e testes de usabilidade são abordados nesta dimensão. A dimensão de *design* de interface também aborda questões de acessibilidade e usabilidade relacionadas para portais de dados, ajudando os tomadores de decisão a tornar o *design* acessível para partes interessadas. Conforme destacado abaixo, uma das maneiras de fazer isso é através do uso de painéis.(MOORE, 2018)

A prototipação ajuda a entender qual o propósito do que foi desenvolvido, o negócio do cliente, a proposição de melhorias e a minimização de riscos e maximização de lucros.(HEBERLE, 2017). Uma das técnicas de design utilizadas nesse trabalho na prototipação é o método 60-30-10, que é "uma regra criada nas áreas de arquitetura e design de interiores. Entretanto, pode ser aplicada a outras áreas como design de interfaces, pois permitem unir cores de forma mais simples. Como boa prática, 60% compreende a cor principal, 30% com a cor secundária, e 10% dedicada à cor de destaque. (CASTILHOS; SANTOS, 2021)

3 TRABALHOS RELACIONADOS

Nesta seção, serão apresentados trabalhos que buscam aprimorar e trazer novas maneiras de realizar a análise de dados dos resultados do exame ENADE, por meio de diversas estratégias, como a criação de aplicações específicas, que permitem uma análise mais eficaz dos dados.

Em (COSTA, 2016), é proposta a criação do *Vis-Scholar*, um *software* que tem como objetivo permitir a visualização de dados do INEP. Os dados são mostrados a partir de um grupo de variáveis, por exemplo: Notas Enade, Notas ENEM. Na parte de análise de dados foi utilizada a técnica de análise de componentes principais - ACP. Após, é feita a exibição de gráficos por estados. Os gráficos contêm resumos de fatores como concorrência, carga horária, notas dos ingressantes, população por aluno matriculados, dentre outros. Como resultado do trabalho houve uma maior contextualização do rendimento da educação em nível estadual. Porém, percebe-se que a ferramenta carece de análises mais específicas, por exemplo: consultas por cidades, *campi*, instituições e etc.

Já no trabalho de (ALVARES; CAMPOS; GOMES, 2015), o autor reforça o quão importante é a análise de grandes volumes de dados, assim como também a importância de visualizações de dados. Com isso, foi visto que no contexto do ENADE os resultados do exame são disponibilizados de forma limitada e que há uma dificuldade na disponibilização de informações.

Com isto é proposta a criação de um *software* que disponibilize os resultados do exame ENADE de uma forma mais detalhada e com uma melhor usabilidade para o usuário. A ferramenta tecnológica utilizada foi a *QlikView 11 Desktop*. O sistema contém várias funcionalidades que, de acordo com a edição, permite análises de regiões, e outras que utilizam a instituição e demais dados para consultas mais específicas, como comparações entre instituições, análises por curso, históricos e resumo por IES. Contudo, o trabalho não permite comparações entre mais de uma edição, e também a comparação de dados históricos entre duas ou mais instituições de ensino.

No trabalho de (ARAÚJO et al., 2019), é abordado o uso de mineração/visualização de dados e sua importância em vários contextos, principalmente seu uso na educação. Com a problemática da dificuldade na análise de dados, foi proposta a criação de uma ferramenta web que permita uma maior exploração de dados do exame chamado ENADE, que, de acordo com os resultados disponíveis pelo INEP, permita uma análise comparativa entre Instituição de Ensino Superior - IES, e os cursos ofertados pela mesma, assim como uma classificação que permita contextualizar o usuário a respeito da análise como um todo. Foram utilizadas técnicas de mineração de dados por meio da linguagem R e *framework Shiny*, a outra técnica foi a Descoberta de Conhecimento em Bases de Dados -KDD, e o algoritmo de mineração de dados utilizado foi o CART. O resultado do processamento dos usos das técnicas abordadas resultam nas funcionalidades como: comparações entre institutos a partir de filtros informados, consultar IES que contém informações como resumo e comparação com os demais institutos, a outra funcionalidade que utiliza da Árvore de Decisão por meio do algoritmo CART, gerando in-

formações complementares de forma clara para pessoas que não têm conhecimento aprofundado no assunto. Por fim, o trabalho possibilitou um formulário para que fosse avaliada a ferramenta, o público alvo da pesquisa foram alunos, professores e coordenadores de cursos. Como resultado da pesquisa tem-se que a aplicação cumpriu o objetivo proposto. Contudo, a ferramenta não permite a comparação de mais de dois cursos ao mesmo tempo, ou demais comparações além de um curso específico com os resultados de todos os demais cursos. Outro ponto é que a ferramenta utiliza dados apenas do ano de 2017, e há uma necessidade de abranger as demais versões do exame.

Já no trabalho de (CUNHA; SALES; SANTOS, 2021), temos a proposta de criação de uma ferramenta que auxilie a análise e com isso uma tomada de decisão por parte de diretores e coordenadores de Institutos de Ensino Superior - IES. Com isso, a ferramenta realiza uma automatização na análise de um determinado curso de Bacharelado em Ciência da Computação - BCC, de acordo com os dados do ENADE disponibilizados pelo INEP. Para a criação do sistema foram usadas a ferramenta *Jupyter Notebook* pra parte de visualização de dados, assim como *Python, Pandas e Matplotlib*, o processo de análise foi baseado no KDD, e a fonte dos dados dos cursos de BCC foi o site da ENADE, com os respectivos anos das edições: 2017, 2014, 2011, 2008, e 2005. O *software* proposto neste trabalho carece de suporte aos demais cursos participantes do exame ENADE, permitindo o acesso de análise somente para o curso de BCC.

Características	COSTA, 2016	ALVARES; CAMPOS; GOMES; 2015	ARAÚJO et al, 2019	CUNHA; SALES SANTOS 2021	SAEN
Permite comparações específicas entre duas universidades(Como cidades, campis, e instituições)?			X		X
Permite uma liberdade na consulta de informações (Opções de filtragens)?	X	X	X		X
Permite comparações entre resultados de edições diferentes (inclusive as edições recentes)?				X	X
Permite analisar áreas diferentes?	X	X	X		X
O trabalho utiliza técnicas de visualização de dados para comparação de informações e exibição ao usuário?	X	X	X		X

Quadro 2. Tabela Comparativa - Fonte: Autor

Na quadro 2 percebe-se que maioria dos trabalhos utilizam de visualização de dados para disponibilização dos dados, além de permitir opções diferentes de filtragem dos dados, no entanto, não exploram tanto as comparações específicas dos resultados, além de não utilizarem das edições mais recentes dos dados, já o SAEN disponibiliza os dados de 2014 a 2021, de todas as áreas de avaliação contempladas no exame ENADE, por meio de uma ferramenta *WEB*.

4 METODOLOGIA

4.1 Coleta de Dados

Os arquivos contendo os resultados das edições do Exame Enade foram coletados no site oficial do INEP, esses arquivos fornecem informações sobre a qualidade do ensino superior e as áreas avaliadas naquela edição. Os arquivos são disponibilizados em Excel e CSV, porém foi utilizado somente o formato Excel devido a facilidade de ler e analisar o conteúdo. Com isto foi baixado e armazenado as edições de 2004 a 2021 na ferramenta Google Drive. O intervalo de anos foi escolhido dessa forma devido ao valor histórico de uma possível análise aprofundada dos dados obtidos.

4.2 Fluxograma da Aplicação

Este trabalho realizou a criação de uma ferramenta web que permite realizar análise dos dados do exame enade de forma mais acessível, que atualmente os dados são disponibilizados de forma individual.

Com isso foi criado o seguinte fluxograma da aplicação:

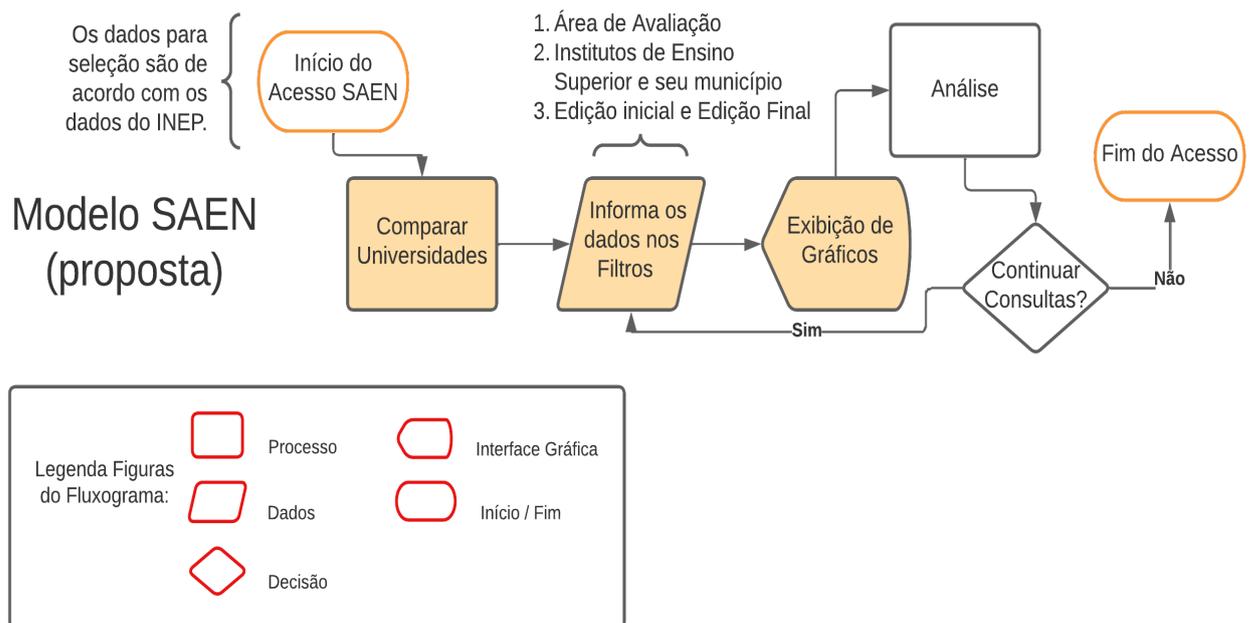


Figura 1 – Fluxograma da Aplicação, Fonte: Autor

A funcionalidade de comparação de universidades tem como objetivo comparar duas universidades de ensino superior, a partir de uma área de ensino em comum, município, edição inicial e final. Em seguida, serão gerados gráficos mostrando os resultados das instituições selecionadas, entre aquele intervalo de edições.

A ordem de escolhas é: 1. Área de Avaliação, 2. Institutos de Ensino Superior. 3. Municípios, 4. Edições. A definição da ordem de seleção ocorreu dessa forma devido à geração automática de informações ao selecionar um campo anterior.

4.3 Coleta de Requisitos

Com o objetivo de serem obtidos requisitos iniciais do sistema foi realizado no dia 08 de fevereiro de 2023 uma reunião com um membro da Pró-Reitoria de Gestão de Pessoas - PROGEP, juntamente com coordenador do curso de Bacharelado em Ciência da Computação do IFCE Campus Aracati no qual foram coletadas informações sobre a necessidade da ferramenta e como poderia ser as funcionalidades. A partir desse encontro, juntamente com o fluxograma da aplicação, foi criada a seguinte lista de requisitos funcionais e não funcionais de acordo com os objetivos do SAEN.

4.3.1 *Requisitos funcionais e não funcionais*

- Requisitos funcionais
 - Consulta de dados: O sistema deve permitir que o usuário escolha quais dados irão ser comparados a partir dos seguintes filtros: Área de avaliação, Nome IES 1, Nome IES 2, Município IES 1, Município IES 2, Edição Inicial e Edição Final.
 - Tipo de Ferramenta: A ferramenta deve ser uma ferramenta WEB.
- Requisitos não funcionais
 - Segurança: O sistema deve garantir a segurança dos dados do servidor, protegendo e evitando acesso não autorizado.
 - Desempenho: O sistema deve ser capaz de realizar consultas e geração de gráficos em tempo hábil, mesmo com um grande número de dados.
 - Usabilidade: A interface deve ser intuitiva e fácil de usar, permitindo que naveguem facilmente e compreendam as informações mostradas nos gráficos
 - Confiabilidade: Os dados do sistema devem ser confiáveis, principalmente por se tratar de notas de instituições de ensino superior.
 - Usabilidade: A ferramenta deverá conter um tutorial.

4.4 Análise de Dados

4.4.1 Processamento de Dados

Os dados foram importados em uma pasta do autor no *Google Drive*, para um rápido acesso na nuvem e segurança em caso de perdas. Com isso foi escolhida a linguagem *Python* e a ferramenta *Google Colab* para ser realizado o processamento de dados de forma *online*. Após criado o arquivo no *Google Colab* na conta do autor, foi feita a leitura a partir do drive do autor (conforme explicado na subseção 4.1 - Coleta de Dados).

No processamento de dados foram utilizados os seguintes critérios para remoção ou ajuste de dados:

- Padronização de colunas;
- Deleção de dados que não contêm informações referentes às notas;
- Deleção de colunas que não irão ser utilizadas no trabalho.

Os critérios estabelecidos na análise dos dados tiveram como foco remover dados irrelevantes para a ferramenta, após a análise foram identificados alguns dados que não estavam completos com as informações necessárias para a ferramenta do SAEN, por exemplo: dados em branco, valores sem utilidade na ferramenta, dados sem notas, e previamente evitar problemas quando realizada a consulta na ferramenta proposta. Além desse tratamento dos dados, foi utilizado outro critério que seria a obrigatoriedade das informações conterem os respectivos campos:

1. Ano.
2. Código da área.
3. Código da IES.
4. Nome da IES.
5. Município do curso.
6. Número de concluintes inscritos.
7. Número de concluintes participantes.
8. Nota bruta FG.
9. Nota padronizada FG.
10. Nota bruta CE.
11. Nota padronizada CE.

12. Conceito Enade contínuo.

13. Conceito Enade faixa.

Outra questão na parte de processamento de dados é que algumas edições tinham padrões diferentes de nomes de colunas, por exemplo: no ano de 2015 tinha a coluna “Área de Enquadramento”, já em 2021 a coluna era “Área de Avaliação”. o mesmo caso ocorre para as colunas “Concluintes Participantes”, “Concluintes Inscritos”, “Município”, entre vários outros tipos de dados.



Figura 2 – Dados Deletados no processamento, fonte: Autor

Na figura 2, temos a quantidade de dados deletados por edição, sendo que no total foram deletados 1913 dados, de acordo com o critério escolhido neste trabalho.

Após a análise de forma individual de cada edição, ficou definido o padrão normalizado nas seguintes colunas:

Ano, Categoria Administrativa, Conceito ENADE (Contínuo), Conceito ENADE (Faixa), Código da IES, Código da Área, Código do Curso, Código do Município, Modalidade de Ensino, Município do Curso, Nome da IES, Nota Bruta - CE, Nota Bruta - FG, Nota Padronizada - CE, Nota Padronizada - FG, Nº de Concluintes, Nº de Concluintes Inscritos, Organização Acadêmica, Sigla da IES, Sigla da UF, Área de Avaliação.

Após todo o processamento de dados, foram exportados os dados padronizados de forma individual, em arquivos de formato CSV. A escolha do formato foi influenciada pela facilidade de ler os dados na linguagem escolhida para inserção de informações, que será explicado com detalhes nas seções seguintes.

4.4.2 Modelagem de Dados

De acordo com os dados obtidos no processamento e após a definição das funcionalidades, foram definidas as seguintes tabelas:

Conforme a Figura 3 (Modelagem de Dados), a tabela INSTITUTO_IES irá armazenar todas as informações específicas do Instituto de Ensino Superior. Já a tabela AREA_AVALIACAO irá armazenar dados específicos da Área de Avaliação. E a tabela EXAME_ENADE irá guardar informações do desempenho de uma instituição em uma determinada área de avaliação.

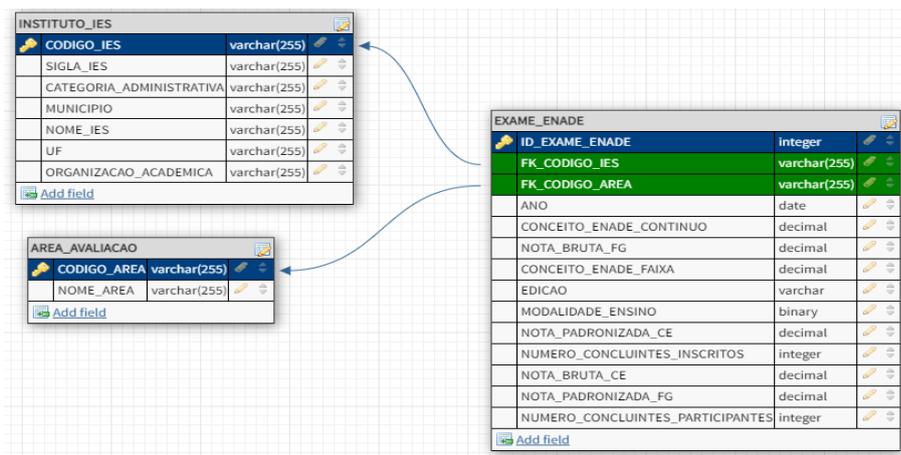


Figura 3 – Modelagem dos Dados, Fonte: Autor

Também foram definidas chaves primárias contendo dados únicos e que contém um nível elevado de importância para o funcionamento da aplicação no banco de dados, sendo elas o CÓDIGO_IES da tabela INSTITUTO_IES, CÓDIGO_AREA da tabela AREA_AVALIACAO e ID_EXAME_ENADE da tabela EXAME_ENADE. A relação entre as tabelas é estabelecida através das chaves estrangeiras FK_CODIGO_IES e FK_CODIGO-AREA da tabela EXAME_ENADE.

4.4.3 Criação do Banco de Dados

O banco de dados escolhido foi o MySQL, devido a sua estrutura relacional por motivo de organização dos dados em tabelas, que podem facilitar no entendimento e na manipulação de informações.

A inserção de dados foi feita através de uma aplicação em Java/Spring. A escolha da linguagem e *framework* para inserção foi devido à facilidade da linguagem na conectividade com um banco de dados, fornecendo um *driver* para conexão.

1. Leitura dos arquivos CSV dos anos de 2014 a 2021. Os arquivos foram lidos de uma pasta local do autor, após o processamento na subseção 4.3.1(Processamento de Dados). A limitação de edições teve como objetivo melhorar o desempenho das análises de informações e no futuro serem implementadas as edições restantes.
2. Inserção de forma individual. utilizado o tratamento de exceções, juntamente com os comandos do *Driver* da linguagem Java: o *rollback* em casos de erros e *commit* caso ocorrido com sucesso.

4.4.4 Criação da API(*Application Programming Interface*)

O desenvolvimento da *API* foi feito utilizando Java como linguagem de programação, juntamente com o *Spring Framework*. Em seguida foram criadas as seguintes opções de consultas:

1. Consultar todas as áreas;
2. Consultar edições por áreas;
3. Consultar todos os municípios;
4. Consultar todos os Institutos de Ensino Superior - IES;
5. Consultar dados por ano e município.

No geral, foram utilizados parâmetros de acordo com os filtros da aplicação para a montagem dos gráficos da aplicação. Para testes da *API* foi utilizado a ferramenta gratuita chamada *Postman*, que permite realizar requisições de vários métodos *HTTP* diferentes. no entanto, Nesta aplicação foi utilizado somente o *GET* para obter os dados. A escolha da ferramenta de teste foi devido a popularidade da ferramenta e facilidade de compartilhar documentos a partir dela.

4.5 Prototipação da Interface

A prototipação da interface foi feita utilizando a ferramenta Figma, que é uma ferramenta de prototipação de interface permitindo a criação da prototipação de alto nível (a mais próxima da versão final). As cores foram escolhidas de acordo com a técnica de *design* chamada 60-30-10, sendo: sendo 60% a cor dominante, 30% a cor secundária, e 10% a cor de destaque.

Desta forma, foi pensado na seguinte paleta de cores:



Figura 4 – Paleta de Cores Sistema, Fonte: Autor

Em seguida foi criado o protótipo de uma *interface* no figma, utilizando da paleta de cores da Figura 7.



Figura 5 – Protótipo - Tela Principal, Fonte: Autor

5 RESULTADOS

5.1 Funcionamento da Ferramenta

A ferramenta contém duas opções de navegação, a primeira seria a funcionalidade "Comparar Universidades" e a segunda seria a página "Sobre", que irá disponibilizar mais detalhes sobre o objetivo e o funcionamento da aplicação. e para facilitar o acesso dos usuários foi desenvolvido um tutorial que contém três demonstrações de como usar a ferramenta da maneira correta.

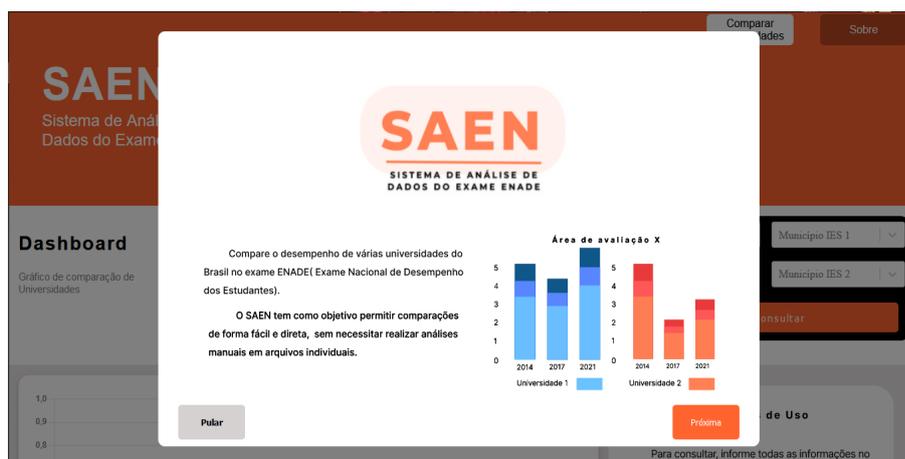


Figura 6 – SAEN - Tela Tutorial, Fonte: Autor

O filtro da aplicação tem 7 campos, sendo eles: Área, Instituição de Ensino Superior 1, Instituição de Ensino Superior 2, Município IES 1, Município IES 2, Edição Inicial e Edição. Existem três gráficos, sendo o primeiro da comparação de universidades, o segundo e o terceiro referentes a quantidade de alunos inscritos.

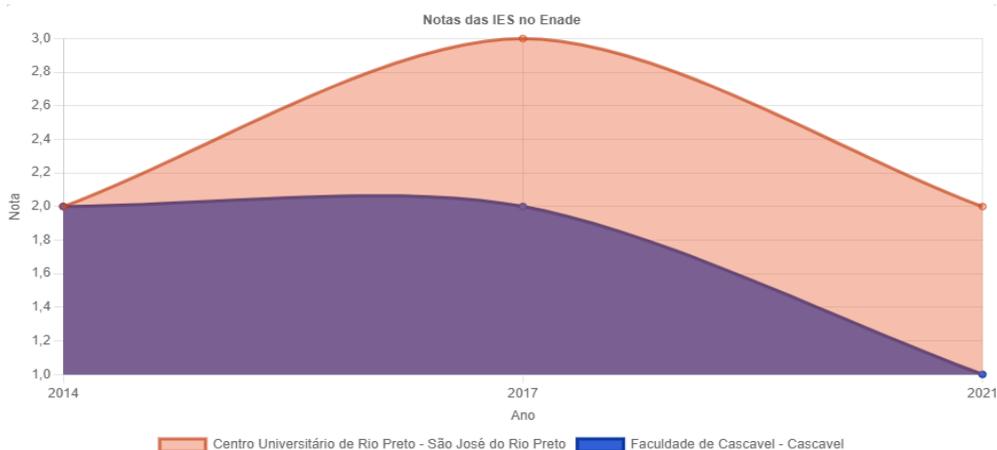


Gráfico 1 - Gráfico de notas das IES, Fonte: Autor

No gráfico 1 são comparados os Institutos de Ensino Superior entre o intervalo de edições selecionado. Já no gráfico 2 mostra os dados entre as edições de alunos inscritos e alunos participantes.

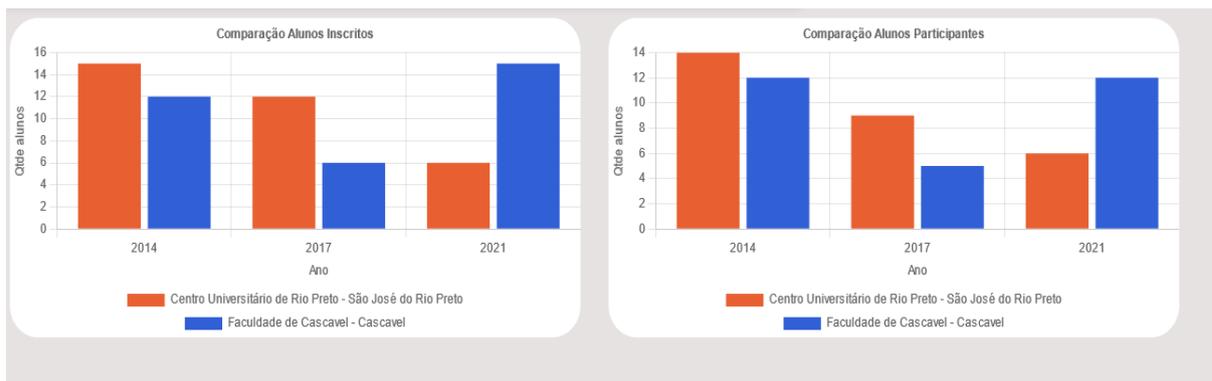


Gráfico 2 - Comparação quantidade alunos inscritos e participantes, Fonte: Autor

A página sobre foi criada com o objetivo de contextualizar a criação da ferramenta e fornecer informações do seu uso. Os tópicos foram separados entre a explicação do que é o ENADE, e o que é o SAEN.



Figura 7 – Página Sobre, Fonte: Autor

5.2 Exemplos de consultas

5.2.1 Comparação duas universidades diferentes

Neste primeiro exemplo, foi selecionado os seguintes dados:

Ciência da Computação | Instituto Federal de Educação, Ciência e Tecnologia... | Aracati

2014 | 2021 | Centro Universitário das Faculdades Metropolitanas... | São Paulo

Consultar

Figura 8 – Filtros do SAEN, Fonte: Autor

Com isso, ao clicar em "Consultar", forma gerados os seguintes gráficos:

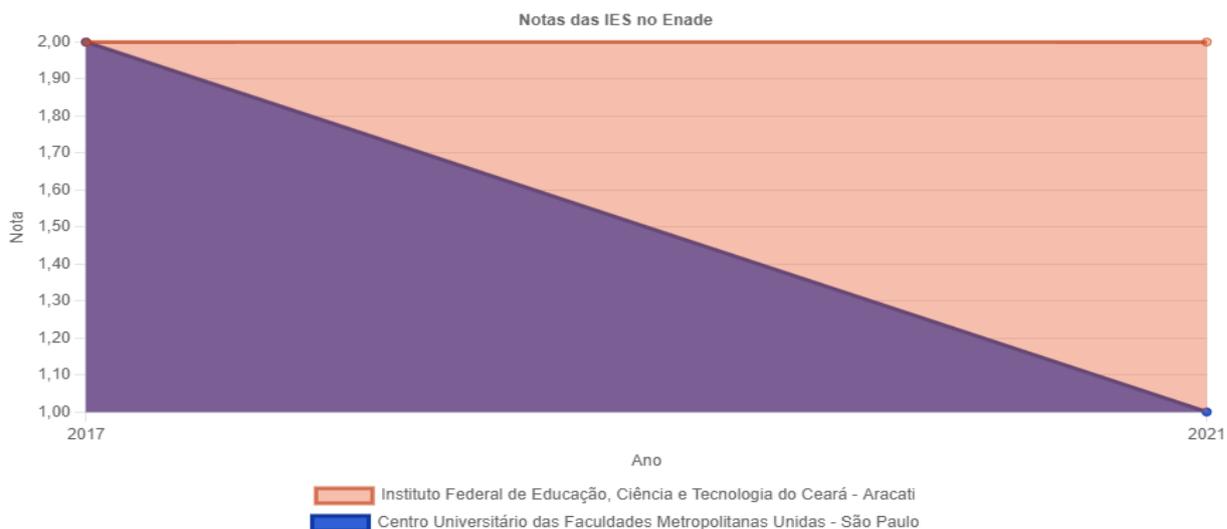


Gráfico 3 - Gráfico de notas das IES exemplo 1, Fonte: Autor

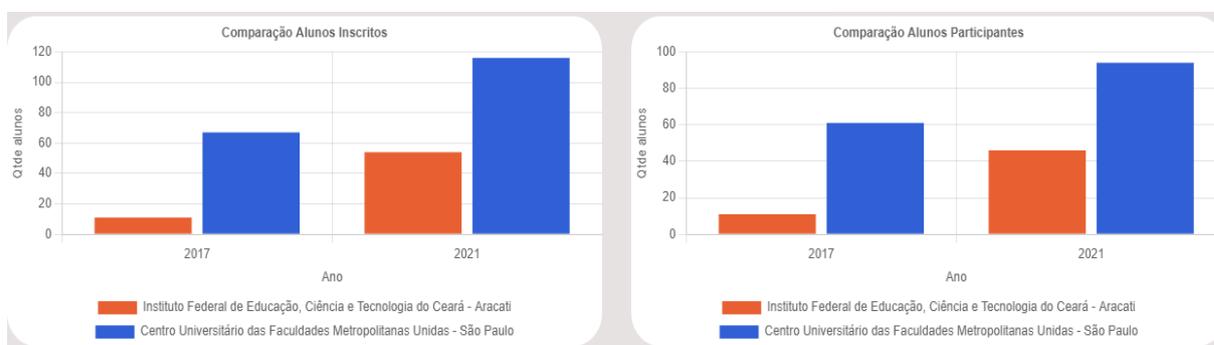


Gráfico 4 - Comparação alunos inscritos e participantes exemplo 1, Fonte: Autor

Nota-se neste exemplo que o Instituto Federal de Educação, Ciência e Tecnologia do Ceará - Aracati obteve a nota 2 no ano de 2017, e permaneceu com a nota 2 no ano de 2021. Já o Centro Universitário das Faculdades Metropolitanas Unidas - São Paulo obteve a nota 2 no ano de 2017, porém obteve um menor rendimento no ano de 2021 comparado a outra IES.

Em relação aos gráficos de alunos inscritos e participantes, nota-se uma grande diferença na quantidade de alunos entre as duas instituições, porém no ano de 2021 a quantidade de alunos da IES de Aracati aumentou bastante comparado a antes.

5.3 Comparação de diferentes *Campis* da mesma universidade

Figura 9 – Filtros exemplo 2, Fonte: Autor

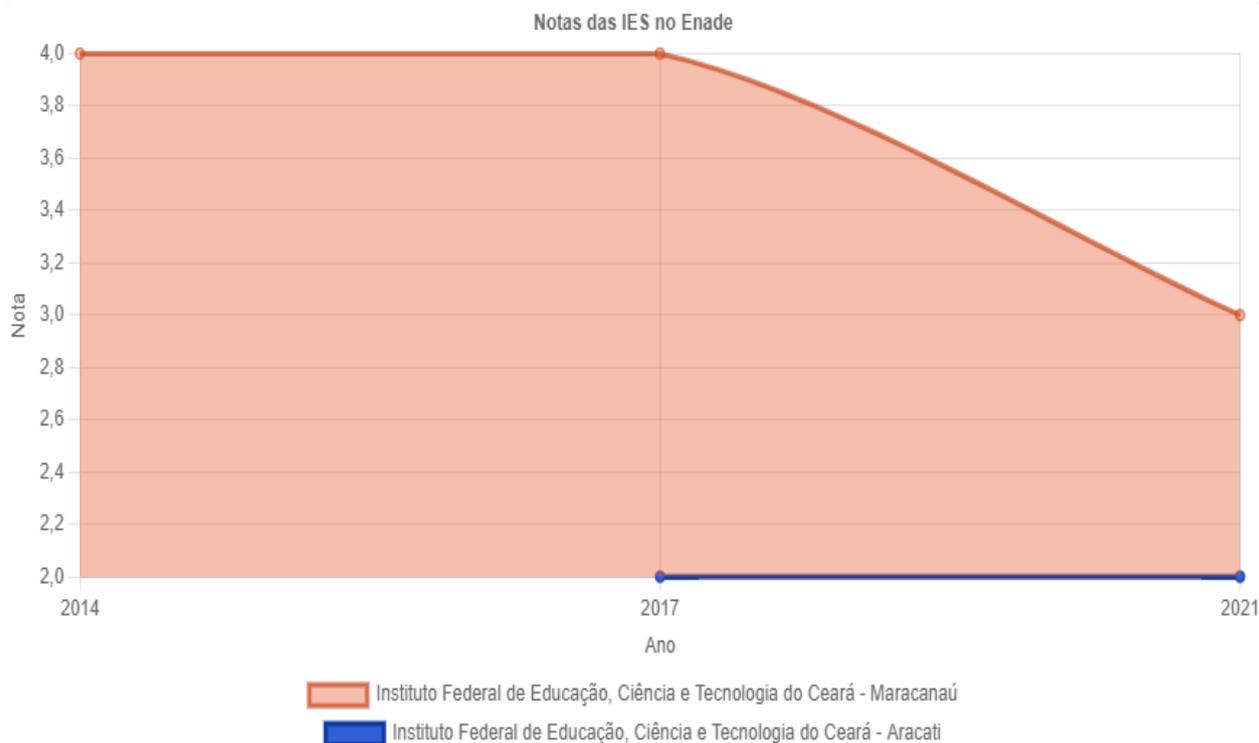


Gráfico 5 - Gráfico de notas das IES exemplo 2, Fonte: Autor

No exemplo 2 teve como objetivo mostrar uma outra maneira de comparação, que seria a comparação de *campis* diferentes da mesma Instituição de Ensino. Neste, foi escolhido o Instituto Federal de Educação, Ciência e Tecnologia do Ceará, com os municípios de Maracanaú e Aracati. Neste há um cenário um pouco diferente do Exemplo 1, pois percebe-se que o campus Aracati não realizou o ENADE no ano de 2014, somente realizou no ano de 2017 e 2021, obtendo o conceito Enade 2 em ambas edições. Já o campus de Maracanaú obteve a nota 4 no ano de 2014 e 2017, já no ano de 2021 obteve a nota 3. Com isso vemos que o campus de Maracanaú obteve um rendimento melhor comparado ao Campus de Aracati nas edições anteriores.

5.4 Pesquisa de Satisfação

Foi criado um formulário no Google Forms com o objetivo de coletarmos informações referente a eficiência da ferramenta, com o título "IFCE Aracati | Pesquisa - SAEN TCC Breno Brito", após foi compartilhado nos principais meios de comunicação do IFCE.

As perguntas foram criadas abordando os principais aspectos da ferramenta, com o objetivo de avaliar os seguintes critérios:

- Produtividade
- Usabilidade
- Velocidade
- Valor Histórico

As perguntas foram separadas em 10 perguntas, sendo 9 objetivas e 1 subjetiva.

A pergunta 1 foi criada com o objetivo de identificar o perfil da pessoa que está respondendo o formulário. As perguntas de 1 a 9 são objetivas, já a 10 é subjetiva e tem como objetivo obter feedbacks gerais ou sugestões para trabalhos futuros. No geral, foram obtidas 52 respostas e para cada pergunta, foram obtidos os seguintes dados estatísticos:

5.4.1 Pergunta 1 - Em qual das opções abaixo você se enquadra?

Das 52 respostas, 33 pessoas que responderam foram de alunos do ensino superior, 3 professores, 5 professores coordenadores, 2 alunos do ensino médio e 9 pessoas com outros perfis.

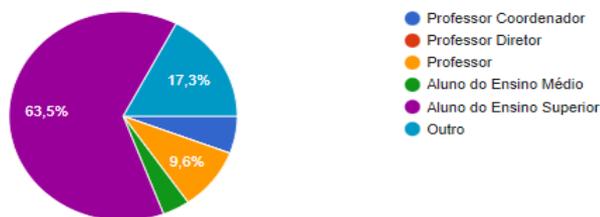


Gráfico 6 - Resultados da Pergunta 1, Fonte: Formulário de pesquisa de Satisfação do Autor

5.4.2 Pergunta 2 - Quão intuitiva é a interface da aplicação para você?

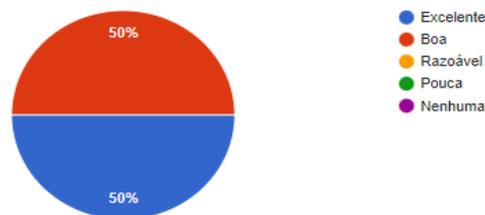


Gráfico 7 - Resultados da pergunta 2, Fonte: Formulário de pesquisa de Satisfação do Autor

26 Pessoas escolheram a opção "Excelente", e 26 escolheram a opção "Boa".

5.4.3 Pergunta 3 - O SAEN é uma ferramenta de fácil uso na realização das consultas?

33 pessoas escolheram a opção "Excelente", 18 escolheram a opção "Boa", e 1 pessoa escolheu a opção "Razoável".

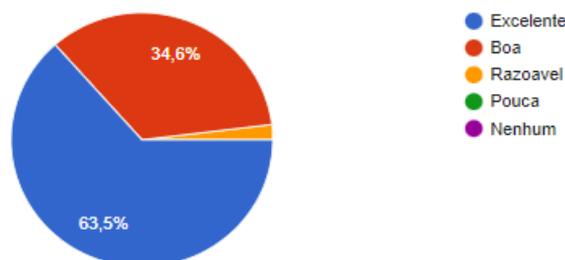


Gráfico 8 - Resultados da pergunta 3, Fonte: Formulário de pesquisa de Satisfação do Autor

5.4.4 Pergunta 4 - Como você avalia a organização e a apresentação dos resultados?

31 pessoas escolheram a opção "Excelente", 15 escolheram a opção "Boa", e 6 escolheram a opção "Razoável".

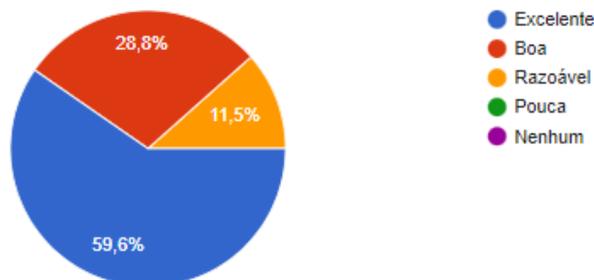


Gráfico 9 - Resultados da pergunta 4, Fonte: Formulário de pesquisa de Satisfação do Autor

5.4.5 Pergunta 5 - Como você avalia a velocidade das respostas?

42 pessoas escolheram a opção "Excelente", e 10 pessoas escolheram a opção "Boa".

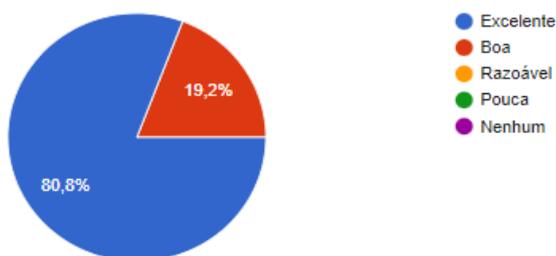


Gráfico 10 - Resultados da pergunta 5, Fonte: Formulário de pesquisa de Satisfação do Autor

5.4.6 Pergunta 6 - Quão satisfeito você está com a clareza e legibilidade da ferramenta?

26 Pessoas escolheram a opção "Excelente", 23 escolheram a opção "Boa", e 3 escolheram a opção "Razoável".

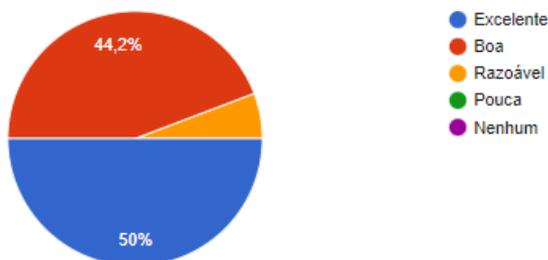


Gráfico 11 - Resultados da pergunta 6, Fonte: Formulário de pesquisa de Satisfação do Autor

5.4.7 Pergunta 7 - Quão bem a ferramenta atende as suas necessidades na comparação de resultados do exame ENADE?

28 escolheram a opção "Excelente", 19 escolheram a opção "Boa", e 5 escolheram a opção "Razoável".

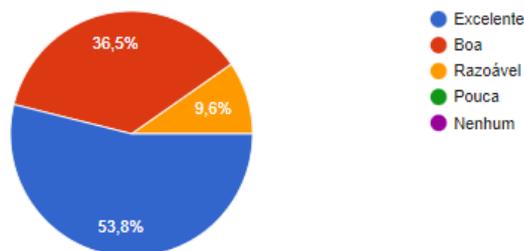


Gráfico 12 - Resultados da pergunta 7, Fonte: Formulário de pesquisa de Satisfação do Autor

5.4.8 Pergunta 8 - Como você avalia a variedade dos filtros(opções de consulta)?

24 pessoas escolheram a opção "Excelente", 24 escolheram a opção "Boa", 2 escolheram "Razoável" e 2 escolheram a opção "Pouca".

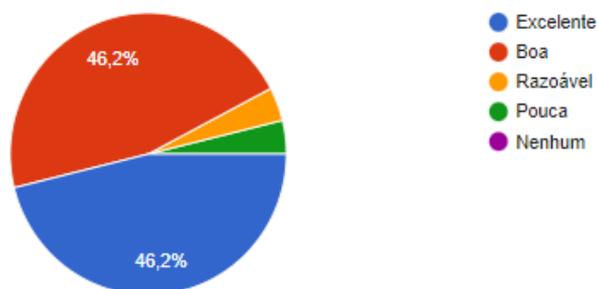


Gráfico 13 - Resultados da pergunta 8, Fonte: Formulário de pesquisa de Satisfação do Autor

5.4.9 Pergunta 9 - Como você avalia a facilidade de uso da ferramenta para realizar as consultas?

30 pessoas escolheram a opção "Excelente", 21 escolheram a opção "Boa", e 1 escolheu a opção "Razoável".

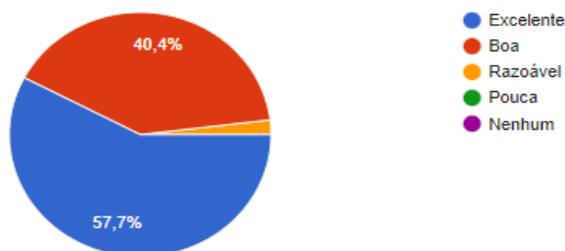


Gráfico 14 - Resultados da pergunta 9, Fonte: Formulário de pesquisa de Satisfação do Autor

5.4.10 Pergunta 10 - Deseja deixar algum comentário sobre a ferramenta? Como sugestões ou opiniões.

Na pergunta 10 obtivemos 17 respostas, as principais sugestões faladas foram:

- Ajustes Gráficos
- Ajustes de Usabilidade

Além disso, foram comentados sobre a necessidade deste trabalho, e o quão eficiente a ferramenta é, e que há um potencial grande.

6 CONCLUSÃO

Este trabalho buscou reforçar a importância de criações de ferramentas na área da educação e demonstrar o quão ágil a tecnologia pode ser em relação a análise de dados. Tendo como princípios a usabilidade, sendo uma ferramenta de fácil acesso a todos usuários por meio de uma aplicação *WEB*. Com foco na área educacional, com acessos para gestores, coordenadores, diretores e professores, assim como também alunos. Além de permitir uma imersão de informações para interesses externos como pais, estudantes do ensino médio e afins, e até mesmo para empresas em caso de doações, investimentos, participações e vínculos de eventos ou bolsas, estágios e projetos extracurriculares, dentre outros.

A ferramenta proposta abordou estratégias diferentes para análises de informações, de acordo com os dados fornecidos pelo INEP. Outro ponto importante é a relação histórica dos dados e as opções de consultas e filtros de edições de 2014 a 2021.

Atualmente a aplicação se encontra hospedada, além de o projeto estar em um repositório no GitHub.

• Links da aplicação:

1. <<https://ui-sistemaanalisedadosenade-tcc.vercel.app/>> (Aplicação SAEN)
2. <<https://github.com/brenobccc/ui-sistemaanalisedadosenade-tcc.git>> (Repositório da Interface do Usuário | FrontEnd SAEN)
3. <<https://github.com/brenobccc/api-enade-analisedados.git>> (Repositório da API | BackEnd SAEN)

Para trabalhos futuros, a ferramenta pode incorporar um módulo didático visando a preparação para o exame ENADE, permitindo verificar quais questões foram mais difíceis em determinadas edições e áreas dos cursos aplicados. Juntamente a isso, visa a implementação de inteligência artificial para identificação de padrões de matérias, conteúdos, questões e permitir visões de preparações adequadas a cada situação entre as áreas incluídas no exame ENADE. Além disso, tem-se como trabalho futuro também a melhoria da usabilidade nas consultas do SAEN, assim como a responsividade para telas de celulares.

REFERÊNCIAS

ALVARES, R. V.; CAMPOS, N. d. S.; GOMES, V. B. Adoção de data discovery para apoio ao processo de análise de dados do enade. In: **CONGRESSO INTERNACIONAL DE INFORMÁTICA EDUCATIVA**. [S.l.: s.n.], 2015. v. 20, p. 480–485.

ARAÚJO, R. A. et al. Análise dos microdados do enade: Proposta de uma ferramenta de exploração utilizando mineração de dados. Universidade Federal de Goiás, 2019.

BAKER, R.; ISOTANI, S.; CARVALHO, A. Mineração de dados educacionais: Oportunidades para o Brasil. **Revista Brasileira de informática na educação**, v. 19, n. 02, p. 03, 2011.

BORCHARDT, G. T. et al. Ferramenta de visualização de dados do censo da educação superior do Inep. In: SBC. **Anais do X Workshop de Computação Aplicada em Governo Eletrônico**. [S.l.], 2022. p. 227–234.

BOSCARIOLI, C.; VITERBO, J.; TEIXEIRA, M. F. Avaliação de aspectos de usabilidade em ferramentas para mineração de dados. In: SBC. **Anais da I Escola Regional de Sistemas de Informação do Rio de Janeiro**. [S.l.], 2014. p. 87–94.

CASTILHOS, L. F.; SANTOS, Y. N. d. Plataforma para organização sistemática de material didático online. 2021.

COSTA, F. G. d. Visualização de dados e sua importância na era do big data. 2017.

COSTA, J. C. A. Vis-scholar: uma metodologia de visualização e análise de dados na educação. Universidade do Vale do Rio dos Sinos, 2016.

CUNHA, R.; SALES, C.; SANTOS, R. Análise automática com os microdados do Enade para melhoria do ensino dos cursos de ciência da computação. In: SBC. **Anais do XXIX Workshop sobre Educação em Computação**. [S.l.], 2021. p. 208–217.

Diário Oficial da União. **Edital nº 37, de 25 de maio de 2023**. 2023. Disponível em: <<https://www.in.gov.br/en/web/dou/-/edital-n-37-de-25-de-maio-de-2023-486214440>>.

EDUCAÇÃO, M. da. **INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira**. 2023. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade>>.

FERREIRA, M. et al. Panorama comparativo da atuação acadêmica e profissional de egressas e egressos de um curso de bacharelado em ciência da computação. In: SBC. **Anais do XXX Workshop sobre Educação em Computação**. [S.l.], 2022. p. 227–238.

HEBERLE, M. Ferramenta para prototipação de interfaces e apoio ao mapeamento de requisitos de sistema. 2017.

IGNÁCIO, S. A. Importância da estatística para o processo de conhecimento e tomada de decisão. **Revista Paranaense de Desenvolvimento-RPD**, n. 118, p. 175–192, 2010.

INEP. **Nota Técnica nº 7/2022/CGCQES/DAES: Metodologia de Cálculo do Conceito Enade 2021**. 2022. <https://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2019/nota_tecnica_n_7_2022_CGCQES_DAES_metodologia_calculo_conceito_enade_2021.pdf>.

International Organization for Standardization. **ISO 9241-11:2018 Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts**. 2018. Disponível em: <<https://www.iso.org/standard/63500.html>>.

KUNZAYILA, N. Uso da visualização de informações para análise de dados educacionais: um estudo de caso sobre a área acadêmica, financeira e científica da Unikiv. Universidade Estadual Paulista (UNESP), 2021.

MATSUBA, D. S.; MATTEDI, A. P. Visualização de dados para extração de conhecimento: um estudo de caso. **AtoZ: novas práticas em informação e conhecimento**, v. 10, n. 2, p. 66–77, 2021.

MOORE, R. L. The role of data analytics in education: Possibilities and limitations. In: **Responsible Analytics and Data Mining in Education**. [S.l.]: Routledge, 2018. p. 101–118.

TEIXEIRA, E. B. A análise de dados na pesquisa científica: importância e desafios em estudos organizacionais. **Desenvolvimento em questão**, v. 1, n. 2, p. 177–201, 2003.

TRASSI, V. A. C. Visualização da informação: análise de ferramentas web para auxílio a tomada de decisão. 2016.