

UNIVERSIDADE FEEVALE

GUSTAVO WILLRICH

PLATAFORMA PARA VISUALIZAÇÃO DOS MICRODADOS DO  
ENEM  
(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo  
2020

GUSTAVO WILLRICH

PLATAFORMA PARA VISUALIZAÇÃO DOS MICRODADOS DO  
ENEM

(Título Provisório)

Anteprojeto de Trabalho de Conclusão de  
Curso, apresentado como requisito parcial  
à obtenção do grau de Bacharel em  
Ciência da Computação pela  
Universidade Feevale

Orientador: Dr. Juliano Varella de Carvalho

Novo Hamburgo  
2020

## RESUMO

A avaliação de indicadores educacionais é um importante instrumento de gestão, permitindo à rede de ensino identificar situações que necessitem de mudanças, incentivos ou aprimoramentos. O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) realiza regularmente diversas avaliações, com o objetivo de monitorar a qualidade do ensino no Brasil, disponibilizando posteriormente os resultados para o público em geral na forma de microdados, permitindo assim a realização de estudos independentes. Uma dessas avaliações é o Exame Nacional do Ensino Médio (ENEM), que visa avaliar o desempenho dos estudantes das instituições de ensino. Embora os microdados do ENEM estejam disponíveis ao público, o grande volume de dados torna o processo de visualização e análise complexo, desestimulando a utilização desses dados no âmbito do setor educacional. Em função disso, este trabalho propõe o desenvolvimento de uma plataforma Web para auxiliar na visualização e análise dos microdados do ENEM, a fim de permitir que a gestão pública, instituições de ensino e pesquisadores da área da educação possam usufruir dessas informações de forma visual e analítica.

Palavras-chave: Visualização de Dados, ENEM, Microdados do ENEM, INEP.

## SUMÁRIO

MOTIVAÇÃO .....	5
OBJETIVOS .....	8
METODOLOGIA .....	9
CRONOGRAMA .....	10
BIBLIOGRAFIA .....	11

## MOTIVAÇÃO

A comunicação de forma visual não é um assunto novo, as imagens têm sido usadas como meio de comunicação mesmo antes da formalização da linguagem escrita e continuam presentes no cotidiano das pessoas, seja no mapa da previsão do tempo demonstrando o deslocamento de tempestades, em placas de trânsito informando sobre uma curva acentuada na estrada ou mesmo em campanhas de marketing incentivando a aquisição de produtos e serviços. Um dos principais motivos para isso é devido a interpretação de informações em imagens ser realizada em paralelo com o sistema de percepção humano, sendo esse muito mais rápido quando comparado com a velocidade na interpretação de informações escritas, pois estas acabam sendo limitadas pelo processo sequencial de leitura (WARD; GRINSTEIN; KEIM, 2015).

No contexto de análise de dados, ao longo dos anos foram criados diversos métodos para representar determinados conjuntos de informações de forma gráfica, alguns métodos se destacam em popularidade e ainda dominam as salas de reuniões, tais como, gráficos de pizza, barras e linhas, mesmo sendo originários do século XVIII. Porém, com o avanço da capacidade tecnológica, tanto da coleta quanto do processamento de dados, atualmente dispõem-se uma vasta gama de recursos que possibilitam visualizar grandes conjuntos de dados de forma eficiente, facilitando o processo de análise. (KIRK, et. al., 2016).

Segundo Knaflic (2019), "A tecnologia permitiu acumular volumes de dados cada vez maiores e há um também crescente desejo associado de entendê-los." Ward, Grinstein e Keim (2015) complementam que em praticamente qualquer domínio, a visualização de dados está se tornando uma ferramenta eficaz para auxiliar a análise e comunicação.

Segundo Kirk et. al. (2016) a visualização de dados é um processo de descoberta, permitindo ao leitor não apenas ver os dados brutos, mas compreendê-los e convertê-los em informação útil. Mazza (2009) reforça que para dar sentido a esses dados, deve-se primeiro utilizar ferramentas e métodos adequados para organizar, processar e representar os dados considerando o contexto que se deseja analisar, assim transformando-os em informação.

Uma das áreas que mais se beneficiam do grande volume de dados existentes atualmente é o ambiente corporativo. A adoção de técnicas como *Data Analytics* e *Business Intelligence* sobre os dados históricos de uma organização permite, por exemplo, identificar o perfil de compra dos clientes, os produtos mais vendidos ou se uma campanha de marketing

obteve o resultado esperado, auxiliando no planejamento e tomada de decisões futuras (SHARDA; DELEN; TURBAN, 2019)

De acordo com Alvarez (2020), atualmente, as escolas também estão passando a adotar conceitos de *Data Analytics* e *Business Intelligence*, sobre os dados educacionais para auxiliar na tomada de decisões

No ambiente escolar, essas decisões podem ocorrer em todos os níveis, desde um professor que percebe a necessidade de fazer a revisão para um grupo específico de alunos que não compreenderam certo conteúdo, chegando a um gestor que decide mudar a estratégia de marketing para atrair novos alunos da vizinhança. (Alvarez, 2020).

Uma das fontes de dados educacionais no Brasil é o próprio INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), que disponibiliza para download diversos arquivos de microdados. Estes arquivos constituem no menor nível de desagregação de dados recolhidos por pesquisas, avaliações e exames, como Censo Escolar, ENADE, ENEM, entre outros (INEP, 2020).

O Exame Nacional do Ensino Médio (ENEM) consiste em uma avaliação aplicada anualmente com o objetivo de mensurar a qualidade do ensino no país e o desempenho individual, tanto de estudantes que completaram o ensino médio quanto do público em geral que esteja interessado em avaliar seus próprios conhecimentos. Atualmente o exame é constituído de uma redação e 180 questões de múltipla escolha, divididas em quatro áreas: linguagens, códigos e suas tecnologias; ciências humanas e suas tecnologias; ciências da natureza e suas tecnologias e matemática e suas tecnologias; além de um questionário socioeconômico que permite que sejam realizados estudos levando em consideração a realidade dos estudantes e sua família (INEP, 2020).

Um dos principais objetivos da avaliação educacional é garantir a qualidade do ensino, porém para os estudantes, a nota obtida no ENEM também colabora para o acesso à educação superior por meio de programas de bolsa de estudos, financiamento estudantil, entre outros programas do MEC (INEP, 2020).

Os microdados relacionados ao ENEM são constituídos de arquivos texto em formato “csv”, onde cada linha representa um participante do exame. Cada participante é identificado por meio de um número de inscrição, não sendo possível identificar o estudante diretamente, para cada participante estão disponíveis diversos dados como respostas do questionário socioeconômico, respostas das questões objetivas, nota da redação, município onde foi realizado a prova, dentre outros (INEP, 2020).

A análise dos microdados disponibilizados pelo INEP não é um processo trivial, Almeida Filho (2014) comenta que embora os microdados estejam divulgados ao público na internet, a grande quantidade de dados e a complexidade para analisá-los impede a ampla difusão do conhecimento no âmbito do setor educacional.

Travitzki (2013) descreve sua experiência sobre a utilização dos microdados do ENEM em seu projeto de pesquisa, “No início das pesquisas tivemos muita dificuldade em acessar os dados, pois não utilizamos nenhum dos programas proprietários (SAS e SPSS) para os quais são fornecidos, nos próprios microdados, arquivos de leitura”. Como alternativa, foi necessário modificar funções em linguagem R utilizadas na análise de dados do IBGE. Alves (2018) também comenta que durante sua pesquisa não foi possível visualizar os microdados do ENEM de 2016 diretamente com editores de texto e planilhas comuns como Bloco de Notas, Microsoft Excel e Libreoffice Calc, pois devido ao grande volume de dados acabavam ocorrendo erros nos softwares. Vale ressaltar que na edição do ENEM de 2016 os microdados possuem cerca de 8 milhões de linhas armazenadas em um único arquivo com tamanho de aproximadamente 5 Gigabytes (INEP, 2020).

Atualmente já existem plataformas que compilam os microdados do ENEM, algumas desenvolvidas por empresas privadas, outras através de trabalhos de graduação. No entanto, foi identificado que essas plataformas tendem a ser direcionadas a ajudar as escolas a localizarem sua posição em um ranking geral, comparar sua posição com outras escolas da região e identificar em quais as áreas do ensino seus alunos sentiram mais dificuldades. Desta forma, existem lacunas que podem ser aprimoradas para melhor aproveitar o volume de informações que os microdados do ENEM têm a oferecer, permitindo que pesquisadores, gestores públicos, dentre outros interessados, realizem análises mais abrangentes sem necessariamente ter vínculo com a uma escola específica.

Dada a dificuldade de analisar os dados do ENEM, este trabalho, portanto, propõe o desenvolvimento de uma plataforma Web para visualização dos microdados do ENEM, realizando uma comparação entre as principais plataformas de visualização existentes, a fim identificar as limitações que possam ser aprimoradas, além de realizar um estudo sobre as técnicas e tecnologias para análise e visualização de dados educacionais. Desta forma, espera-se que a plataforma possa contribuir para uma melhor compreensão sobre os fatores que influenciam a qualidade do ensino no Brasil, e permita que não apenas escolas, mas também pesquisadores e a gestão pública possam usufruir destas informações.



## OBJETIVOS

### Objetivo geral

Desenvolver uma plataforma Web que auxilie a visualização e análise dos microdados do ENEM, a fim de permitir que a gestão pública, instituições de ensino e pesquisadores da área da educação extraiam informações relevantes.

### Objetivos específicos

- Identificar bibliografias sobre estudos baseados nos microdados do ENEM.
- Selecionar métodos e técnicas de visualização de dados.
- Identificar plataformas existentes para visualização de dados do ENEM.
- Comparar a plataforma desenvolvida com plataformas pré-existentes.
- Compor diferentes visualizações a partir dos microdados do ENEM.
- Avaliar a plataforma desenvolvida.

## METODOLOGIA

Este trabalho pode ser considerado como pesquisa de natureza aplicada, pois tem como propósito desenvolver uma plataforma Web para visualização e análise dos microdados do ENEM, com o objetivo de permitir que a gestão pública, instituições de ensino e pesquisadores da área da educação extraiam informações relevantes que permitam avanços em suas áreas de atuação.

Os procedimentos técnicos adotados serão pesquisa bibliográfica e pesquisa experimental. Inicialmente será feito o levantamento bibliográfico em trabalhos de conclusão de curso, teses de mestrado e doutorado, pesquisas científicas, artigos, relatórios técnicos e livros sobre pesquisas relacionadas ao ENEM e seus microdados, a fim de identificar os objetivos principais dos usuários e quais os tipos de análise são mais comuns.

Posteriormente, será feito o levantamento e comparação entre características das ferramentas de visualização dos microdados do ENEM encontradas, identificando pontos a serem aprimorados, conforme as necessidades dos usuários que foram identificados na etapa anterior. Posteriormente pretende-se realizar um estudo sobre técnicas de visualização de dados que possam ser incorporadas na plataforma, permitindo a criação de gráficos e visualizações de dados dinâmicas adequadas a área da educação, estimulando a investigação por parte dos usuários.

Após o levantamento do referencial teórico e pesquisa experimental, será realizado uma análise sobre a estrutura dos microdados do ENEM, identificando as diferenças existentes entre cada edição, a fim de selecionar e criar uma base de dados íntegra. Também será realizado o levantamento da infraestrutura necessária para hospedar a plataforma de forma estável.

No Trabalho de Conclusão II (TCII) será realizada a configuração da infraestrutura e o desenvolvimento da plataforma utilizando as informações obtidas durante a elaboração do Trabalho de Conclusão I.

Pretende-se disponibilizar o acesso à plataforma para alguns voluntários da área de ensino com o perfil a ser definido, para fazer uma avaliação baseada em suas próprias experiências e interesses sobre a educação no Brasil. A avaliação tem o objetivo de verificar se a plataforma auxilia ou não aos propósitos previstos, tanto em termos de usabilidade, quanto na qualidade e utilidade das informações apresentadas.

Sendo assim, este trabalho compromete-se a completar o objetivo geral e todos os objetivos específicos apresentados, respondendo a seguinte questão de pesquisa: “Uma plataforma para a visualização e análise dos microdados do ENEM pode auxiliar profissionais que têm necessidade de investigar a qualidade do ensino brasileiro?”

## CRONOGRAMA

### Trabalho de Conclusão I

Etapa	Meses			
	Ago	Set	Out	Nov
Desenvolvimento do anteprojeto	■	■		
Levantamento bibliográfico sobre pesquisas que utilizam os microdados ENEM		■	■	
Análise da estrutura dos microdados do ENEM		■	■	
Comparação características das ferramentas existentes para visualização de dados do ENEM			■	■
Levantamento dos requisitos de infraestrutura para a plataforma			■	■
Desenvolvimento da escrita do Trabalho de Conclusão 1		■	■	■

### Trabalho de Conclusão II

Etapa	Meses			
	Mar	Abr	Mai	Jun
Investigação sobre bibliotecas de visualização.	■			
Criação e configuração da plataforma na Web	■	■	■	
Avaliação da plataforma				■
Desenvolvimento da escrita do Trabalho de Conclusão 2	■	■	■	■

## BIBLIOGRAFIA

ALMEIDA FILHO, Álvaro Cavalcanti de. **Modelo de mensuração do desempenho dos institutos federais: uma análise a partir de microdados**, 2014.

ALVAREZ, Luciana. **Análise de dados ganha espaço na gestão escolar**. Revista Educação, v. 242, 2017. Disponível em: <<https://revistaeducacao.com.br/2017/09/19/analise-de-dados-ganha-espaco-na-gestao-escolar/>>. Acesso em: 19 ago. 2020.

ALVES, Rafael Damiani. **Predição do desempenho da redação do ENEM utilizando técnicas de mineração de dados**. Trabalho de Conclusão de Curso (Graduação) - Curso de Tecnologias da Informação e Comunicação, Universidade Federal de Santa Catarina, Araranguá. 2018.

COSTA, Felipe Garcia da. **Visualização de dados e sua importância na era do Big Data**. Trabalho de Conclusão de Curso (Graduação) - Curso de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre. 2017.

INEP. **ENEM**. Disponível em: <<http://portal.inep.gov.br/enem>>. Acesso em: 12 ago. 2020.

KIRK, Andy; TIMMS, Simon; RININSLAND, Andrew; TELLER, Swizec. **Data Visualization: Representing Information on Modern Web**. Livery Place: Packt Publishing Ltd., 2016.

KNAFLIC, Cole Nussbaumer. **Storytelling com dados: Um guia sobre visualização**. Rio de Janeiro, RJ: Alta Books. 2019.

MAZZA, Riccardo. **Introduction to information visualization**. Springer Science & Business Media, 2009.

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business Intelligence e Análise de Dados para Gestão do Negócio-4**. Porto Alegre, RS: Bookman Editora. 2019.

TRAVITZKI, Rodrigo. **ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar**. 2013. Tese de Doutorado. Universidade de São Paulo.

WARD, Matthew O.; GRINSTEIN, Georges; KEIM, Daniel. **Interactive Data Visualization Foundations Techniques and Applications**. 2nd Edition. Boca Raton, Florida: CRC Press, 2015