

# EduMap: Uma Ferramenta de Analítica Acadêmica para o Mapeamento Geográfico da Origem dos Estudantes

Emanuel Túrmina Torres  
emanueltt1000@gmail.com  
Universidade Federal de Santa Catarina  
Joinville, Santa Catarina, Brasil

Benjamin Grando Moreira  
benjamin.grando@ufsc.br  
Universidade Federal de Santa Catarina  
Joinville, Santa Catarina, Brasil

## ABSTRACT

This article presents the development and application of an academic analytics tool called EduMap, focused on the mapping and geographic analysis of university students. The tool enables data visualization to facilitate the identification of geographic aspects related to the students' origins, aiming to promote data-driven decision-making. An educational analysis is presented based on the proposed tool.

## KEYWORDS

Academic Analytics, Educational Data Analysis, Geographic Mapping, Data Visualization Tool

## 1 INTRODUÇÃO

Analítica Acadêmica (do inglês Academic Analytics) é a expressão utilizada para descrever uso de macrodados originados de sistemas de informação das instituições de ensino para a tomada de decisão administrativa ou gerencial [1]. Diferente de Learning Analytics ou da Mineração de dados Educacionais, a Analítica Acadêmica é mais voltada à questões administrativas do que o enfoque no processo de ensino-aprendizagem.

No contexto da Educational data mining (EDM), a integração de diversas abordagens visa aprimorar a compreensão e predição de padrões, sendo necessário considerar não apenas as técnicas analíticas, mas também os tipos de dados envolvidos nesse processo [2]. No entanto, para uma compreensão abrangente e precisa, é crucial considerar também dados demográficos, como etnia, nível de escolaridade, idade e gênero. Essas variáveis são obtidas por meio de pesquisas, registros acadêmicos e outras fontes externas e são consideradas como possíveis preditores para determinar o êxito acadêmico dos estudantes [3].

Um projeto de Analítica Acadêmica depende de ferramentas tecnológicas que auxiliem na tomada de decisões com base em dados institucionais [1]. Este trabalho apresenta o desenvolvimento de um recurso de Analítica Acadêmica, com foco na distribuição geográfica de origem dos alunos de cursos de graduação. Opções de filtragem estão disponíveis para direcionamento das observações realizadas. O entendimento da origem dos alunos do curso pode auxiliar na identificação de políticas de auxílio estudantil, aspectos de acolhimento dos estudantes e também em estratégias de divulgação do curso.

## 2 MATERIAIS E MÉTODOS

A ferramenta inicial desenvolvida neste estudo é alimentada com dados obtidos via exportação do Sistema de Controle Acadêmico

da Universidade Federal de Santa Catarina (UFSC). Os dados mostrados são relacionados com os 8 cursos de graduação ofertados no Departamento de Engenharias da Mobilidade, campus Joinville.

O relatório de dados cadastrais é composto por informações sobre o curso, situação acadêmica, dados sociodemográficos, e histórico escolar dos estudantes, delineando um panorama do perfil discente e fornecendo substrato para análise preditiva. A Tabela 1 apresenta um dicionário dos dados presentes no relatório. Esse conhecimento se prova importante uma vez que novos indicadores podem ser criados para expandir o escopo do relatório através da manipulação desses dados.

No contexto da análise acadêmica da instituição, utilizam-se índices como o Índice de Aproveitamento Acumulado (IAA) e o Índice de Aproveitamento Probatório (IAP) para avaliar o desempenho dos estudantes. Tais índices contribuem para o entendimento de diferentes aspectos de performance, proporcionando uma análise das diferentes trajetórias tomadas pelos estudantes. O IAA é uma métrica que reflete o desempenho ao longo do curso. O IAP avalia o desempenho com base unicamente na eficiência das aprovações.

Dentro do âmbito sociodemográfico, informações sobre sexo, raça/cor, data de nascimento, estado de origem e semestre de ingresso fornecem uma base para o entendimento das diversas facetas que compõem o corpo discente. A análise do perfil dos estudantes permite a compreensão da diversidade e das dinâmicas sociais presentes dentro da instituição acadêmica [4].

## 3 TRATAMENTO DOS DADOS

Para realização do tratamento dos dados foram seguidos três passos principais: a limpeza, manipulação e a segmentação dos dados. Cada um desses processos é descrito a seguir.

A etapa inicial na limpeza dos dados envolveu a anonimização dos estudantes, removendo dos relatórios atributos relacionados à sua identificação, como nome, sobrenome, endereço de e-mail e ID na instituição.

Em seguida, foram tratados os valores ausentes. No relatório de dados cadastrais, os valores faltantes dos índices numéricos foram substituídos pela média, enquanto os índices de categoria utilizaram o valor mais frequente ou relacional.

Para concluir, foram excluídos índices irrelevantes para a análise com base em sua redundância, baixa variação grupal ou falta de influência discernível nos parâmetros de interesse. Para os dados cadastrais, por exemplo, índices como *Curso* e *nomeCurso* classificam o mesmo tipo de informação, *Nacionalidade* e *estadoCivil* possuem um valor que representa mais de 98% de seus casos e, finalizando, o nome da escola em que um estudante concluiu o ensino médio, representado por *EscolaSG*, não contribui significativamente com a análise devido à sua ampla quantidade de opções.

**Tabela 1: Dicionário de dados dos atributos cadastrais mais relevantes para o trabalho**

Termo	Descrição
Curso	Código numérico do curso.
nomeCurso	Nome completo do curso frequentado pelo estudante.
Situacao	Situação acadêmica atual do estudante (ex: regular, trancado).
semestreSituacao	Semestre correspondente à situação acadêmica registrada.
Sexo	Gênero do estudante.
racaCor	Categoria de raça ou cor do estudante.
dataNascimento	Data de nascimento do estudante.
estadoCivil	Estado civil do estudante.
Nacionalidade	Nacionalidade do estudante.
Naturalidade	Local de nascimento do estudante.
anoSemestreIngresso	Ano e semestre de ingresso do estudante na instituição.
formaIngresso	Forma de ingresso na instituição (ex: vestibular, ENEM, transferência).
categoriaIngresso	Categoria de ingresso do estudante (ex: ampla concorrência, cotas).
ensinoPublico?	Indica se o estudante veio de escola pública (sim ou não).
IAP	Índice de aproveitamento baseado em aprovações.
IAA	Índice de aproveitamento acadêmico acumulado.
ClassificacaoVestibular	Classificação do estudante no processo seletivo de ingresso.
OrdemClassificacao	Ordem de classificação dentro da categoria de ingresso.
EscolaSG	Nome da escola de ensino médio do estudante.
MunicipioSG	Município onde se localiza a escola de ensino médio do estudante.
UFSG	Unidade federativa da escola de ensino médio do estudante.
AnoConclusãoSG	Ano de conclusão do ensino médio pelo estudante.

Após a limpeza, os dados foram submetidos a uma série de manipulações. Essa etapa envolveu a criação de novos índices, a concatenação de informação de alguns e alteração direta do significado de outros.

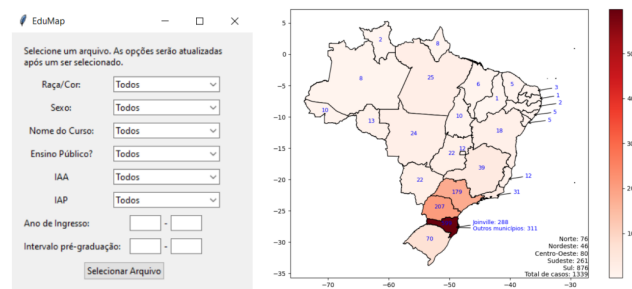
Foram criados índices como *idadeIngresso* e *reprovou?*, os quais substituíram índices originais, permitindo seu uso como variáveis contínuas nos modelos. Em seguida, *UFSG* foi adaptado junto de *MunicipioSG* para discernir quais estudantes de SC são de Joinville<sup>1</sup> e quais são de outros municípios. Esse passo se fez importante pois há um elevado número de estudantes oriundos dessa cidade, conforme será apresentado posteriormente. Além disso, foi renomeado para *origemRegiao* e unificou os estados em suas regiões. Por fim, *anoSemestreIngresso* foi modificado para representar apenas o ano de ingresso, sem a presença do semestre.

#### 4 RESULTADO - SOFTWARE PARA MAPEAMENTO E ANÁLISE GEOGRÁFICA

O desenvolvimento de uma ferramenta para mapeamento e análise geográfica, nomeada EduMap, iniciou-se com a identificação da necessidade de um meio eficiente para visualizar e interpretar os conjuntos de dados cadastrais gerados pelo sistema acadêmico. Para atender a esse requisito, optou-se por uma abordagem modular em Python, integrando tecnologias de front-end para visualização e bibliotecas de back-end para processamento e cálculo.

A EduMap permite ao usuário selecionar e carregar diretamente o arquivo contendo o relatório, sem a necessidade de manipulação prévia do documento. Esta característica é particularmente eficaz

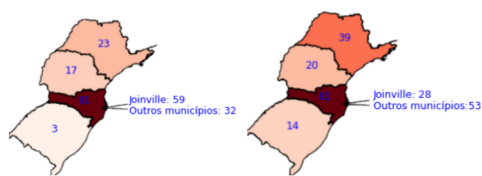
<sup>1</sup>Não foi possível omitir o município para blind review devido a natureza da aplicação e que será evidenciada em figuras posteriores.

**Figura 1: Ferramenta EduMap: menu da ferramenta e exemplo de resultado**

devido ao fato de que o relatório possui um padrão estabelecido, garantindo assim a consistência e a precisão na análise das informações importadas. Este recurso não apenas melhora a usabilidade e a flexibilidade da ferramenta, mas também assegura que os usuários possam realizar análises customizadas com base em dados atualizados.

A concepção da interface gráfica da ferramenta foi planejada para garantir uma experiência de usuário fluida e acessível. Conforme Figura 1, os usuários são capazes de navegar pelo sistema e selecionar variáveis de interesse. A escolha de concentrar-se nessas características foi motivada pela concepção de relevância que tais dados têm na compreensão da análise.

Ao aplicar os filtros selecionados, o programa carrega um mapa temático (heatmap) do Brasil. Este mapa mostra a distribuição dos estudantes, colorindo os estados com intensidades diferentes com



**Figura 2: Distinção regional dos cursos de Engenharia Civil de Infraestrutura (a esquerda) e Engenharia Naval (a direita), para a população branca**

base no número de ocorrências. O mapa também apresenta numericamente o volume de casos por estado, com adendo especial para Santa Catarina, onde os estudantes, devido a seu alto número, são divididos entre residentes de Joinville e residentes de outros municípios. Além disso, é exibida a distribuição por região e o número total de casos para a combinação de características selecionadas. O usuário pode refazer o processo de filtragem e visualização quantas vezes desejar, permitindo uma análise flexível de diferentes conjuntos.

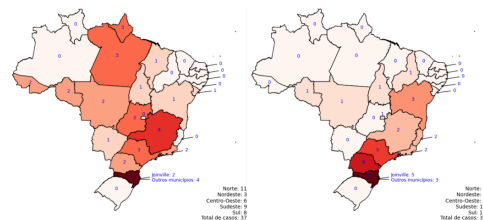
O processo iterativo de desenvolvimento enfatizou a prototipagem rápida e o teste contínuo de novas funcionalidades, assegurando que a ferramenta atenda às necessidades de análise de padrões demográficos, suportando, assim, o planejamento e a pesquisa no contexto educacional do presente trabalho. O código-fonte da ferramenta está disponível no repositório do Github<sup>2</sup>.

#### 4.1 Exemplo de Análise Exploratória

A seguir são apresentados resultados da análise exploratória realizada através da ferramenta apresentada e que ajudam a ilustrar o uso da ferramenta. A análise foi dividida em duas etapas e revelou padrões na distribuição geográfica e demográfica dos estudantes. Inicialmente, é apresentada a análise que visou identificar a origem regional dos cursos com base na raça e gênero dos estudantes, e, em seguida, a que aborda a distribuição de cada grupo com base no IAA.

Ao separar as amostras por raça, os grupos de população branca e amarela apresentaram focos de origem nos estados de Santa Catarina (SC), Paraná (PR) e São Paulo (SP). Além disso, a concentração em cada estado varia entre os cursos. Para os cursos de Engenharia Mecatrônica e de Engenharia Aeroespacial, por exemplo, há uma alta quantidade de estudantes oriundos do PR. Já nos cursos de Engenharia Naval e Engenharia Civil de Infraestrutura, a origem dos estudantes está predominantemente em SC e SP, com uma distinção notável em SC, conforme Figura 2: a maior parte dos alunos de Engenharia Civil de Infraestrutura são da cidade de Joinville, enquanto os de Engenharia Naval vêm de outras cidades do estado.

Para a população parda e preta, observa-se uma distribuição mais equilibrada e distinta por curso em todo o território nacional. O curso de Engenharia Mecatrônica apresenta uma distribuição uniforme, enquanto os cursos de Engenharia Automotiva e Engenharia Aeroespacial têm concentração nos estados litorâneos. Já os cursos de Engenharia de Transporte e Logística, Engenharia Civil e Engenharia Naval têm foco nos estados do sul e no Pará. Um



**Figura 3: Distinção regional entre Engenharia Mecatrônica (a esquerda) e Engenharia Automotiva (a direita) para população parda**

exemplo dessa distribuição está presente na Figura 3, na qual as populações dos cursos de Engenharia Mecatrônica e Engenharia Automotiva são comparadas.

As análises das outras combinações de raça, curso e gênero seguiram padrões semelhantes aos discutidos, mostrando distribuições comparáveis em termos regionais e demográficos. Embora não tenham sido abordadas detalhadamente neste artigo, essas combinações também oferecem informações valiosas sobre a diversidade e a distribuição dos estudantes.

## 5 CONSIDERAÇÕES FINAIS

Este artigo apresentou o desenvolvimento e aplicação de uma ferramenta chamada EduMap, projetada para realizar análises acadêmicas com foco no mapeamento e análise geográfica de estudantes universitários. A ferramenta visa facilitar a identificação de padrões geográficos relacionados à origem dos alunos e seu impacto no desempenho acadêmico. Com isso, o estudo contribui para a promoção da tomada de decisões baseadas em dados. Além disso, o artigo inclui uma análise educacional que exemplifica a utilidade da ferramenta no contexto acadêmico.

As análises realizadas com o EduMap demonstram o potencial da ferramenta como um recurso para instituições de ensino superior. Como trabalhos futuros, recomenda-se a ampliação do estudo para incluir outras variáveis que possam influenciar o sucesso acadêmico, bem como a adaptação da ferramenta para diferentes contextos educacionais, potencializando seu uso em uma variedade de instituições e cursos. A ferramenta deve incorporar outras formas de análise, propondo dashboards, bem como incorporar dados de desempenho acadêmico.

## REFERÊNCIAS

- [1] Andrea Cristina Filatro. *Data Science na Educação: Presencial, a Distância e Corporativa: Presencial, a Distância e Corporativa*. Saraiva Educação SA, 2020.
- [2] Moises Riestra-González, Maria del Puerto Paule-Ruiz, and Francisco Ortín. Massive lms log data analysis for the early prediction of course-agnostic student performance. *Computers & Education*, 163:104108, 2021.
- [3] R. Woodman. Investigation of factors that influence student retention and success rate on open university courses in the east anglia region, 2001.
- [4] Ponni Iyer Ivan Reid Govardhan Wankhede Mary Thornton, Patricia Bricheno and Roger Green. Diversity and social integration on higher education campuses in india and the uk: student and staff perspectives. *Research in Post-Compulsory Education*, 15(2):159–176, 2010. doi: 10.1080/13596741003790682.

<sup>2</sup>Repositório Github do projeto: [https://github.com/emanuelth/Heatmap\\_AE](https://github.com/emanuelth/Heatmap_AE)