

Data Warehouse Educacional: Uma Visão sobre Evasão no Ensino Superior

G.A.S.Santos^{1,2}, L.Tarrataca¹, D.Haddad¹, A.Bordignon³,
D.Brandão¹, K.T.Belloze¹

¹ Programa de Pós-graduação em Ciências da Computação – CEFET/RJ – Brasil

² Superintendência de Tecnologia da Informação – UFF – Brasil

³ Instituto da Matemática – UFF – Brasil



{gassantos, alexb}@id.uff.br

{luis.tarrataca, diego.haddad, diego.brandao, kele.belloze}@cefet-rj.br

1. Introdução

A evasão é uma situação que ocorre quando os estudantes ocupam vagas e se desassociam das universidades sem concluir o curso em que se matricularam.

- Dados do Censo de Educação Superior no Brasil, em 2016, mostraram que das 10,6 milhões de vagas oferecidas nos cursos de graduação, 26% foram oriundas de evasão.
- A Organização para a Cooperação e Desenvolvimento Econômico (OCDE) revela que o custo médio anual de um estudante do ensino público superior no Brasil, em 2013, foi de US\$13.539,90.

Este trabalho descreve a implementação de um *Data Warehouse* Educacional (EDW - *Educational Data Warehouse*), a fim de fornecer um banco de dados analítico capaz de apresentar análises integradas em uma IFES.

2. Data Warehouse Educacional

Na maioria dos projetos de DW, o desafio consiste em planejar de maneira eficiente o desenvolvimento geral do sistema. A Figura 1 apresenta o fluxo de dados para o processo que incorpora o contexto informacional necessário para a construção do EDW.

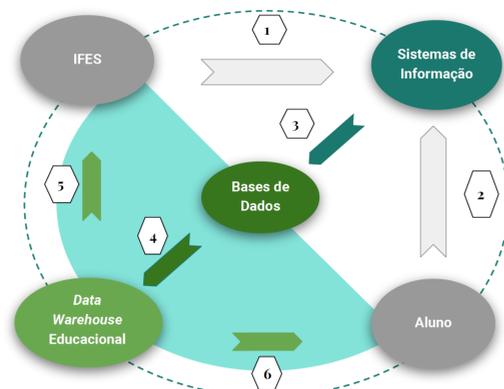


Figura 1. Fluxo de Dados para o *Data Warehouse* Educacional.

Neste processo, estão representadas as seguintes etapas:

- (1) e (2) lidam com as interações entre estudantes e universidade;
- (3), as informações geradas são armazenadas em bases de dados;
- (4) representa o processo de *Data Warehousing* e a base de dados analítica;
- (5) fornece as informações analíticas à tomada de decisão;
- (6) visa conceder recomendações baseadas no EDW.

O desenvolvimento do EDW, focado no tema evasão, foi baseado nos requisitos funcionais apresentados na Tabela 1.

Tabela 1. Requisitos funcionais para o EDW.

Item	Descrição
01	Analisar a evasão com relação ao desempenho acadêmico
02	Analisar a evasão com base no perfil dos estudantes
03	Analisar a evasão considerando a localidade do curso

Para atender os requisitos apresentados na Tabela 1, o conjunto de bases de dados escolhido é referente aos sistemas: idUFF, PIBIC, Sisbol e Monitoria.

3. Estudo de Caso com foco na Evasão

- O conjunto de dados produzido pelo EDW contém informações sobre os estudantes como as notas do ENEM, histórico acadêmico, ações afirmativas e dados sociodemográficos.
- O EDW contém aproximadamente 80.000 alunos dos 106 cursos de graduação presencial da UFF, dentre os anos de 2005 até 2018.

Na Figura 2, é possível identificar a distribuição dos alunos por localidade da instituição, considerando o percentual referente à faixa de coeficiente de rendimento (CR) do estudante e localidade. É apresentado, também, o percentual de CR total da Instituição, o qual mostra que 37% dos alunos apresenta CR abaixo de 4,0.

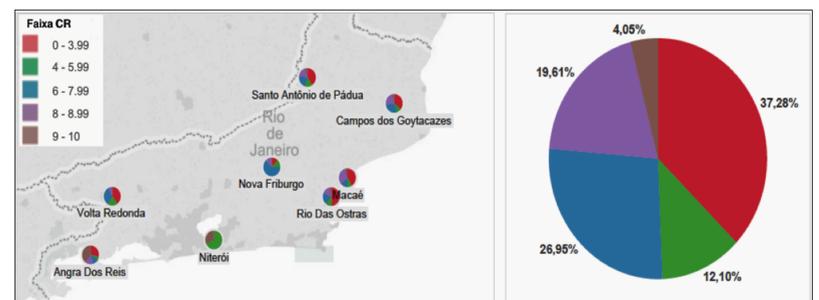


Figura 2. Distribuição do rendimento dos alunos por localidade e total.

Durante a análise exploratória dos dados, algumas perspectivas foram observadas:

- O conjunto de dados continha uma representação de classes desbalanceadas, com 76% de estudantes evadidos e 24% graduados, conforme pode ser visto na Tabela 2.

Tabela 2. Mediana dos atributos baseada nos perfis "Evadido" e "Graduado".

Classe	Qtd.	Idade	SemestreFinal	CR	CargaHoraria
Evadidos	9852	25	1º	3,4	240
Graduados	3117	24	2º	8,3	3199

- Pode-se identificar por meio do histograma apresentado na Figura 3, uma frequência maior da amostra de alunos com baixo desempenho nas regiões de CR variando de '0' até '3', sendo boa parte desta frequência de 34% constituída por alunos evadidos, em geral, ocorrendo esta evasão no primeiro ano de ingresso.

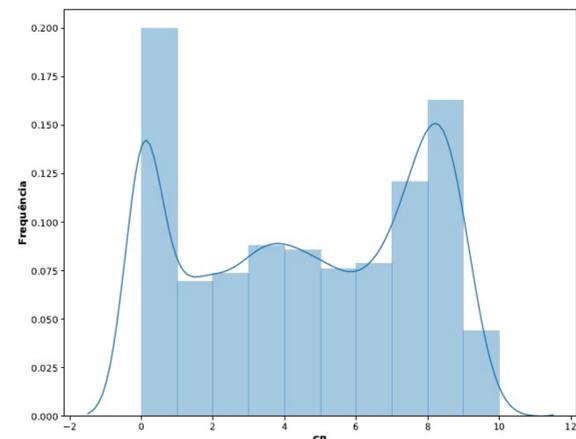


Figura 3. Histograma de CR de toda a amostra, incluindo graduados e evadidos.

4. Considerações Finais

- A proposta mostrou-se como solução capaz de auxiliar a gestão acadêmica na identificação de padrões que impactam na evasão.
- Por meio das análises realizadas, foi percebido que as notas do ENEM têm uma correlação com CR e continuidade do aluno, assim como a carga horária cursada e o tempo de permanência do aluno.
- Evidenciou-se também que a informação escolar progressiva implicará em uma melhor percepção sobre o perfil do aluno de graduação.
- Para trabalhos futuros, serão utilizadas técnicas de aprendizado de máquina para selecionar os atributos mais relevante para representar um aluno e classificá-lo sob perfil de risco de evasão.