

**Universidade Federal do Agreste de Pernambuco**

# Projeto de Pesquisa em Mineração de Dados Educacionais

Orientadores: Rene Pereira de Gusmão e Cleonides Silva Dias Gusmão

Estudantes: Tayane Cibely Batista Rodrigues e José Elias De Lima

## Sobre as bases de dados usadas no Projeto

Bases de Dados do Censo Escolar de 2017 e do SAEB 2017

[https://download.inep.gov.br/dados\\_abertos/microdados\\_censo\\_escolar\\_2017.zip](https://download.inep.gov.br/dados_abertos/microdados_censo_escolar_2017.zip)  
[https://download.inep.gov.br/microdados/microdados\\_saeb\\_2017.zip](https://download.inep.gov.br/microdados/microdados_saeb_2017.zip)

Foram utilizados na Pesquisa, até o momento dos microdados do SAEB:

- TS\_ALUNO\_3EM\_ESC.csv
- TS\_ESCOLA.csv

Em relação ao censo escolar, foi utilizada a base de dado:

- microdado\_ed\_basica\_2017.csv

A partir delas foram utilizados filtros para a obtenção dos dados da Paraíba e Pernambuco, gerando 6 novas bases.

### **SAEB**

- dados\_paraiba\_saeb\_alunos.csv
- dados\_pernambuco\_saeb\_alunos.csv
- dados\_paraiba\_saeb\_escolas.csv
- dados\_pernambuco\_saeb\_escolas.csv

## **CENSO**

- dados\_paraiba\_censo.csv
- dados\_pernambuco\_censo.csv

Para compreensão de como as bases originais estavam organizadas entre as unidades federativas, utilizou-se do seguinte site do IBGE:

<https://www.ibge.gov.br/explica/codigos-dos-municipios.php>

IMPORTANTE: O site esclareceu não só como identificar cada unidade federativa, como os respectivos identificadores dos municípios.

Dados complementares:

github do projeto:

<https://github.com/TayaneCibely/ProjetoPesquisa.git>

Próximo passo:

- Exclusão de variáveis que não agregam valor a pesquisa.
- Merge entre as bases de dados
- Utilização de algoritmos de aprendizagem de máquina.