

**CIn - Centro de Informática da UFPE**

**Ciência dos Dados - IF706**

**Proposta de Projeto**

Gabriel Henrique (ghds) e Anderson Mota (ams9)

Recife - PE / 2017.2

## **Introdução**

A educação é um ponto cada vez mais mencionado como imprescindível para a evolução tanto individual como no contexto da sociedade de modo geral. Diante disto, desejamos trabalhar com dados voltados à educação.

Por questão de escopo, optamos por trabalhar com dados voltados ao ensino médio, principalmente por conta das recentes polêmicas envolvendo o Exame Nacional do Ensino Médio (ENEM), nas quais muitos dos estudantes que prestaram o exame clamam que o nível das questões estava incompatível com o ensino das escolas públicas e somente acessível aos estudantes da rede particular.

## **Objetivos**

Nesse contexto, desejamos investigar as escolas brasileiras e verificar se as diferenças estruturais e em qualidade de ensino são realmente tão impactantes entre as escolas públicas e privadas tanto no aspecto nacional, quanto no regional.

Iremos também tentar classificar as escolas de acordo com seu nível de rendimento no ENEM. Com isso, desejamos investigar os atributos que são comuns a escolas de alto, médio e baixo rendimento para tentar entender quais deles são mais importantes, de forma que uma determinada escola consiga evoluir mais facilmente.

## **Metodologia**

Para tanto, vamos utilizar as ferramentas introduzidas na disciplina para tentar obter respostas para nossos questionamentos através de dados públicos existentes na web, além disso, também estamos investigando obter dados via uma API já existente voltada para educação.

Inicialmente iremos nos preocupar com o processo de extração e pré-processamento dos dados obtidos, de forma a consolidá-los para que seja possível uma análise exploratória inicial. Com isto, poderemos detectar se iremos

conseguir responder nossos questionamentos através desse conjunto de dados, se sim, começaremos a testar nossas hipóteses até chegar na visualização e análise dos resultados. Caso contrário, tentaremos obter novos dados que possuam as informações necessárias para elucidar nossas dúvidas ou em último caso, tentar buscar outra linha de questionamento que possa ser interessante a partir dos dados.