

Projeto de Ciencia dos Dados

Gabriel Henrique Daniel da Silva, Anderson Mota

Recife, Pernambuco

Abstract

Nosso projeto visa investigar dados provenientes da area de educacao, visando entender melhor os atributos que fazem uma escola ser melhor do que outra utilizando como metrica os resultados do ENEM (Exame Nacional do Ensino Medio).

1. Introducao

A educacao e um ponto cada vez mais mencionado como imprescindivel para a evolucao tanto individual como no contexto da sociedade de modo geral. Diante disto, desejamos trabalhar com dados voltados a educacao. Por
5 questao de escopo, optamos por trabalhar com dados voltados ao ensino medio, principalmente por conta das recentes polmicas envolvendo o Exame Nacional do Ensino Medio (ENEM), nas quais muitos dos estudantes que prestaram o exame clamam que o nivel das questes estava incompativel com o ensino das escolas pblicas e somente acessivel aos estudantes da rede particular.

10 2. Objetivos

Nesse contexto, desejamos investigar as escolas brasileiras e verificar se as diferenas estruturais e em qualidade de ensino so realmente to impactantes entre as escolas publicas e privadas. Por questo de escopo, iremos focar nas escolas existentes na cidade de Olinda. Iremos tambm tentar classificar as escolas de

[☆]Fully documented templates are available in the elsarticle package on CTAN.

15 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 investir de forma mais inteligente para evoluir de forma mais simples.

3. Metodologia

20 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. Essa API possui dados sobre diversos
atributos existentes numa escola como por exemplo, se possui biblioteca, acesso
25 a internet, quadras esportivas, além disso também há dados sobre quantidade
de funcionários, alunos, computadores e diversos outros. Também teremos dados
provenientes de uma planilha disponibilizada pelo MEC com as notas do ENEM
das escolas brasileiras. Após obter esses dados e pré-processá-los, iremos partir
para uma análise preliminar dos dados, focada em entender melhor o conjunto
30 de dados. Posteriormente iremos aplicar aprendizagem de máquina para verificar
quais atributos de fato possuem maior importância na qualidade da escola.