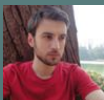


Gerando Insights Acionáveis



Fabio Ocaña Vieira



Gustavo Vitral Arbex

Marcelo Rigotto Stachuk



Tibor Zequini Boglár



Como melhorar a educação básica?

Encontrando insights em exemplos de sucesso!

Mas como encontrar os insights que transformam a educação?



← ATRAVÉS DE DADOS
EXEMPLOS DE SUCESSO →

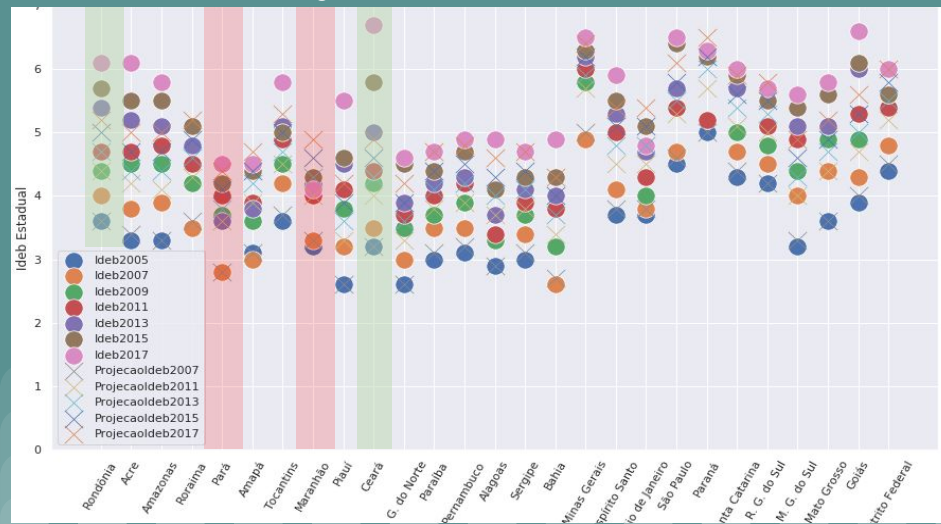


Macro

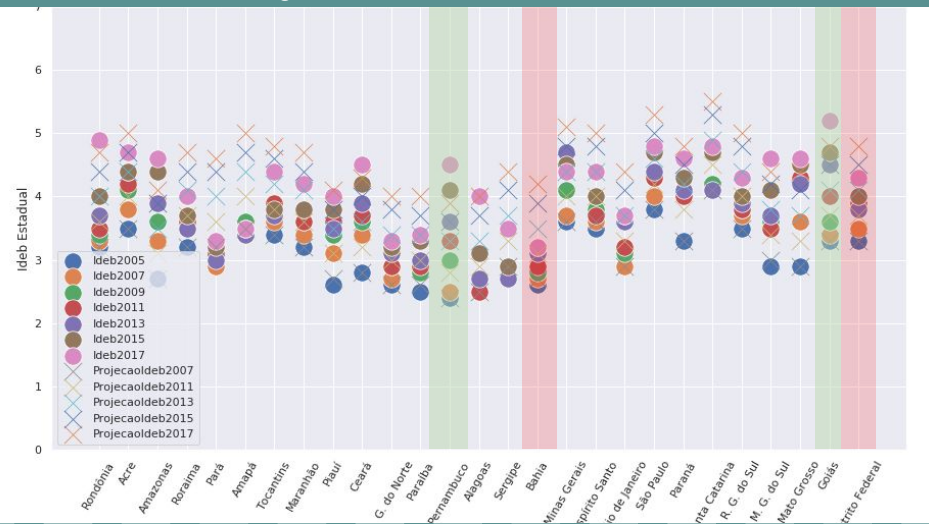
Estados que melhoram sua educação com o tempo podem ser vistos como casos de sucesso!



Progressão do Ideb - Anos Iniciais



Progressão do Ideb - Anos Finais



Políticas Públicas de Sucesso

Goiás e Ceará

CORREIO BRAZILIENSE

Como o Ceará tem garantido liderança em qualidade de ensino?

GAZETA DO POVO

Por que o Ceará têm as melhores escolas públicas do Brasil?

FOLHA DE S. PAULO

★ ★ ★

Goiás inova na sala de aula e vira referência em educação no país

Rede estadual cria material didático que mira as dificuldades dos estudantes

Como atuar:



Parcerias entre governo e 3º setor, implementando políticas públicas que já deram certo.

Como não atuar:

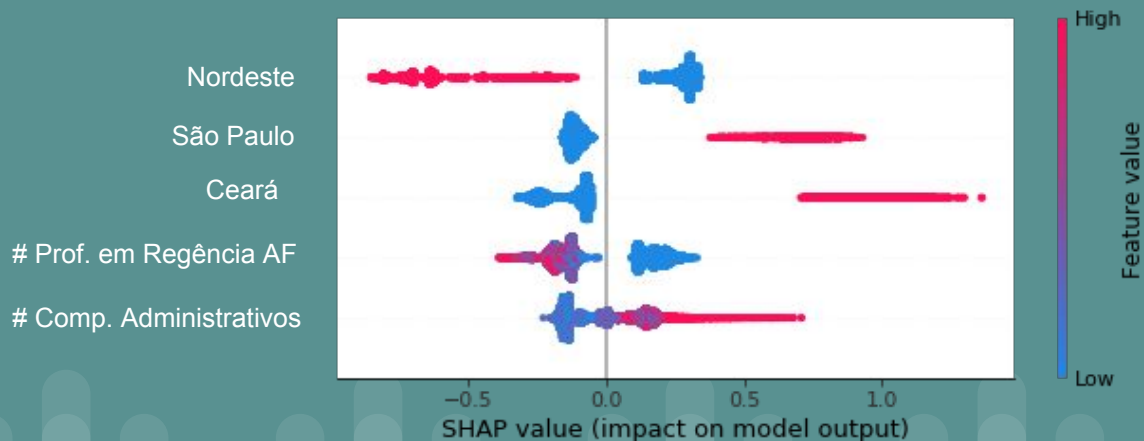


Propondo medidas socioeconômicas de âmbito federal, como aumento da renda e da qualidade de vida.

Mas tiramos isso somente de análises regionais?

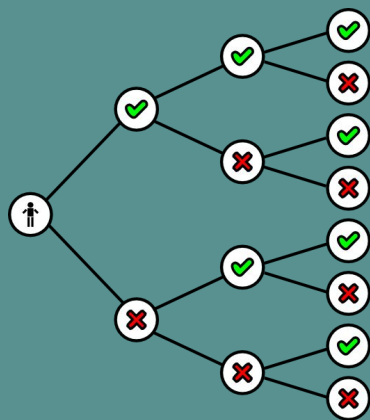
A partir de algoritmos de **classificação** e **explicabilidade**, demonstramos que variáveis regionais são importantes para a classificar uma escola como estar ou não estar entre as 25% melhores no Ideb.

Exemplos de Features que classificam uma escola como sendo boa



Micro

Quais são as semelhanças entre as escolas de melhor IDEB? (Top 25%)



Utilizamos dois ensaios diferentes para determinar essas semelhanças, ambas a partir de métodos de classificação utilizando **decision trees**.

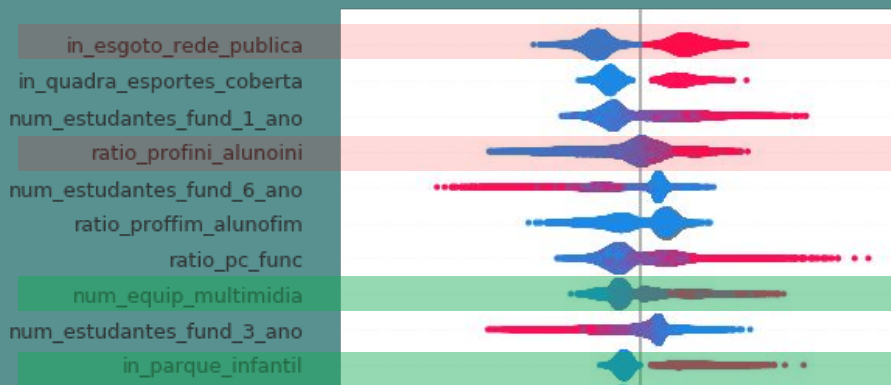
Ensaio 1: A partir de todo conjunto de escolas, gera-se um target binário determinando quais escolas pertencem ao Top 25% no IDEB.

Ensaio 2: Seleciona-se um subconjunto A de escolas, a partir de análises prévias de desigualdade social e notas no IDEB, e novamente cria-se um target binário determinando quais escolas estão 10% acima da média de A.

Alguns Insights

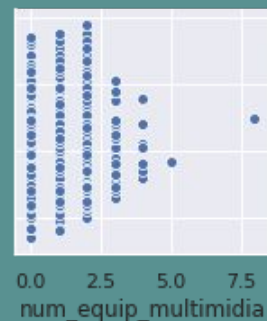
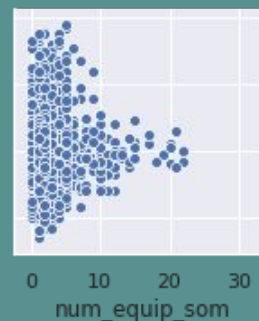
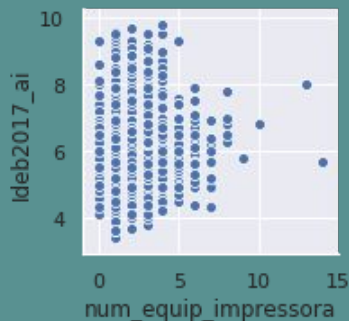
Insights óbvios

1. Rede de esgoto
2. Relação professor / aluno
3. Internet



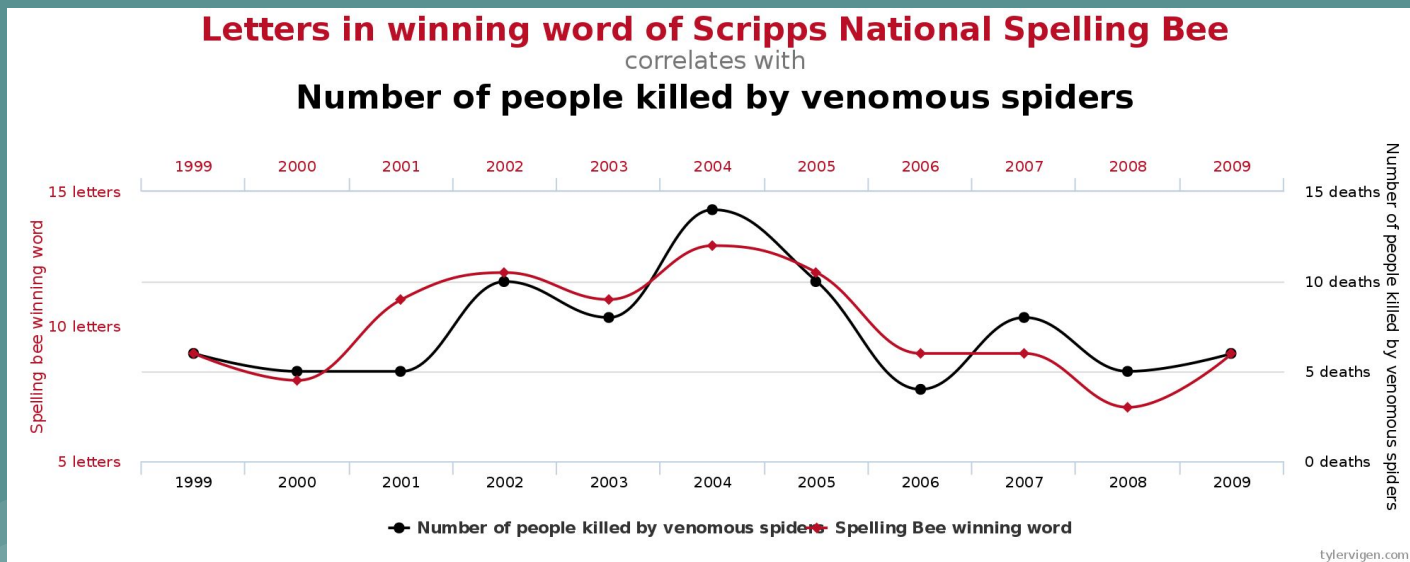
Insights não óbvios

1. Parquinho
2. Equipamentos multimídia



Causalidade não implica Correlação

- ❑ Às vezes as features que representam escolas com IDEB alto não necessariamente causam o IDEB alto;
- ❑ Pode acontecer que escolas com IDEB alto apresentem variáveis parecidas; mas isso não necessariamente é a causa do alto IDEB delas.



Quais as ferramentas para as conclusões

❑ Método de tratamento de dados:

- ❑ Relacionar tabelas para relacionar desempenho com features das escolas;
- ❑ Tratamento de dados categóricos com One Hot Encoding;
- ❑ Criação de novas features que possam indicar bom desempenho:
 - ❑ Razão entre #Professores e #Alunos;
 - ❑ Razão entre #Alunos e #Turmas;
 - ❑ Razão entre #Computadores e #Alunos.



❑ Algoritmos de Aprendizado:

- ❑ Algoritmos de aprendizado extraem informações do conjunto de dados, porém nem sempre são explicáveis.



❑ Algoritmos de Explicabilidade:

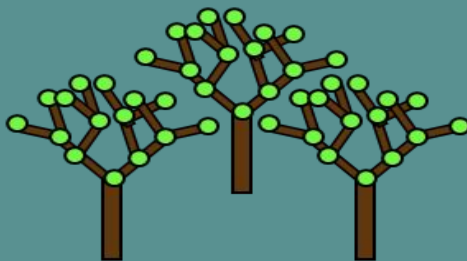
- ❑ Utilizando algoritmos de explicabilidade que atuam junto com os de aprendizado podemos explicar a relevância de cada feature, gerando insights de atuáveis.



Classificação

Decision Trees Classifiers

- ❑ Algoritmo básico para classificação;
- ❑ Seu método de funcionamento envolve a divisão do espaço em regiões conforme as variáveis apresentadas;
- ❑ Foi utilizado um K-Fold em que $K = 10$ para garantir a acurácia.

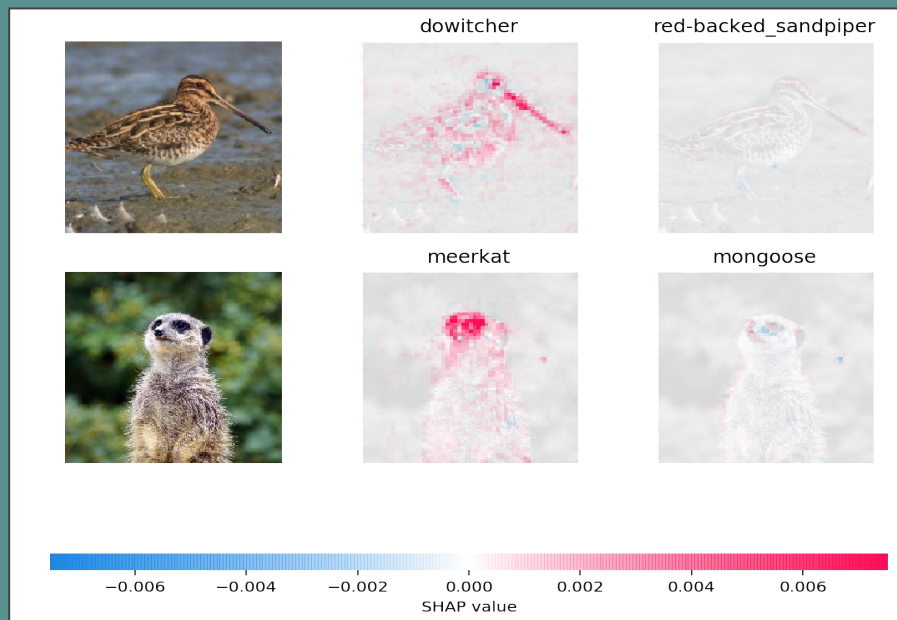


XGBoost

- ❑ Importante algoritmo de Boosting, amplamente utilizado em competições do Kaggle;
- ❑ Baseado em Árvores de Decisão;
- ❑ Aqui foi treinado com Test-Split de 50% obtendo acurácia de aproximadamente 80% em todos casos.

Shap (SHapley Additive exPlanations)

Algoritmo **state-of-the-art** de explicação de modelos complexos



Fontes do Artigo:

<http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions>

<https://github.com/slundberg/shap>

Como atuar?

Como já citado, esperamos que atuações ocorram em parceria com o 3o setor e com o setor público
Porém é necessário criar um modelo para medir a efetividade dos insights para atuar em larga escala

Qual atuação é melhor: Construir uma Quadra ou uma Biblioteca



Como resultado dos algoritmos de classificação, a Feature que representa a existência de quadra na escola é mais importante para a nota do IDEB que se ter uma biblioteca. Sendo assim, fazemos 2 controles um que a compra deveria melhorar o desempenho e outro que não. Portanto se a implantação das quadras derem realmente resultados positivos e o da biblioteca neutros nos livramos do efeito placebo (escola ir bem a curto prazo só por estarem investindo nela) e comprovamos que quadras ajudam no desempenho estudantil. Assim o próximo passo seria de implantar essa política em maior escala.

Conclusões

Existem duas frentes de atuação:

- ❑ **Nível Macro:** Através de Políticas Públicas;
- ❑ **Nível Micro:** utilizar as features mais importantes na análise a fim de criar modelo para ser implementado em um conjunto de escolas.

A partir de análise do nível macro:

- ❑ Encontrar estados que definimos serem outliers; estes apresentam políticas públicas de educação específicas. Assim, a partir de parcerias com governos e o 3º setor, pode-se expandir essas políticas à outros estados.
- ❑ Aumentando a granularidade definimos escolas de sucesso, e com ajuda de algoritmos de aprendizado e interpretabilidade encontramos características em comum entre estas escolas. Por fim definimos um método para verificar se as características de fato influenciam na educação, obtendo uma linha de ação baseado em dados para melhorar o ensino público.

Agradecimentos

Gostaríamos de agradecer, primeiramente, à equipe da **BCG Gamma** pela oportunidade de trabalhar com dados reais em uma problema de importância ao progresso de uma nação.

Também segue o nosso agradecimento à **Fundação Lemann** que busca a melhora da educação pública no Brasil com diversos projetos, acreditando nos sonhos e realizações dos brasileiros.

