

Big data e educação

Advanced analytics como ferramenta de transformação pública

Henrique Sinatura, Tamires Vilela

Julho de 2019

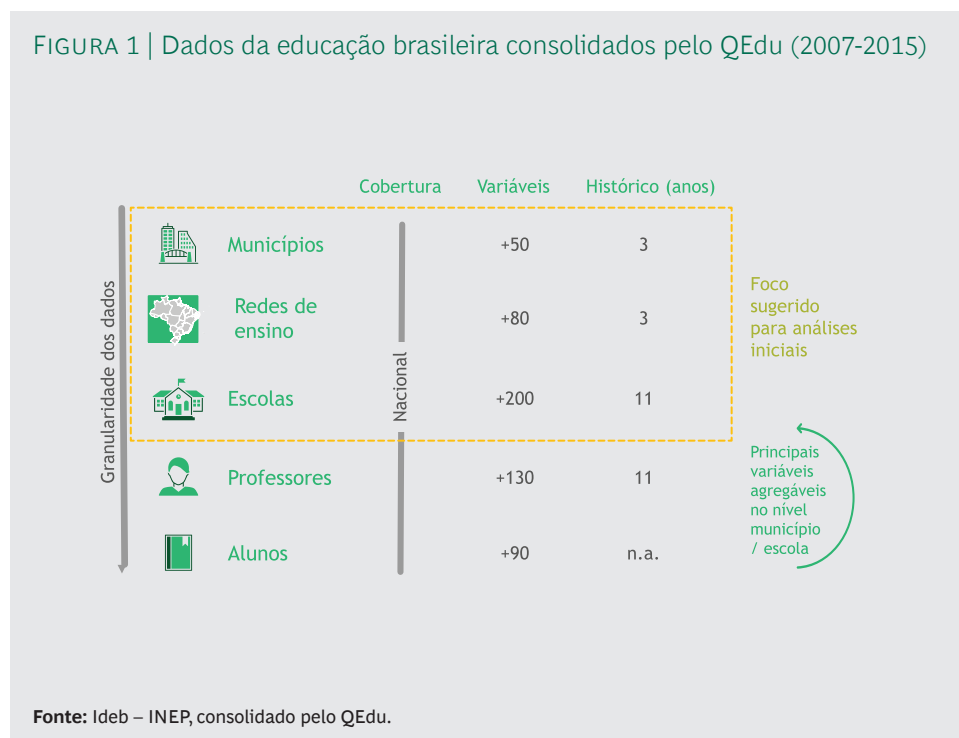
AINTELIGÊNCIA ARTIFICIAL (IA) já é uma realidade no setor privado. Gestores e executivos de todas as indústrias já estão olhando para a IA como uma fonte de vantagem competitiva. Em pesquisa realizada pelo BCG em parceria com o MIT Sloan, mais de 70% dos executivos disseram acreditar que a IA terá papel central e crescente em suas companhias (*Putting Artificial Intelligence to Work*, BCG e MIT Sloan).

Esse novo contexto não é exclusividade apenas do setor privado. Os avanços tecnológicos dos últimos anos facilitaram a coleta, o armazenamento e o processamento de dados, acumulando uma quantidade de informação nunca antes vista também no setor público, onde acreditamos haver oportunidade real para usar tais dados em toda sua capacidade. Apesar da maioria das organizações já trabalharem com um histórico de dados considerável, grande parte não possui uma cultura *data-driven*, tampouco a infraestrutura necessária para trabalhar com tanta informação. A capacidade de capitalizar essas tendências transformando os dados em ativos (ou *data assets*) mostra-se fator determinante para melhorar a entrega dos serviços públicos.

Considerando que o acesso a dados educacionais não é mais uma barreira, a educação no Brasil pode ser vista como uma das áreas do setor público que mais apresenta um campo fértil para o uso de *advanced analytics*. Por exemplo, o QEdu, iniciativa da Fundação Lemann, consolida informações publicadas pelo Governo Federal (por meio do INEP) em múltiplos níveis de granularidade, constituindo uma ampla base de dados e com histórico de até 11 anos (Figura 1). O desafio é utilizar toda a capacidade tecnológica existente para extrair o máximo de valor dos dados disponíveis.

“Gerir uma rede de ensino ou formular políticas públicas com acesso limitado a dados é como pilotar um avião medindo a direção do vento pela copa das árvores. No Brasil, precisamos não só de melhores e mais instrumentos para guiar nosso trabalho, mas também de mais dirigentes capazes de interpretar e tomar boas decisões com base em dados e evidências. Hoje, o QEdu funciona como um painel de controle para ajudar gestores e formuladores de política pública em educação”. Tamires Vilela, Fundação Lemann

FIGURA 1 | Dados da educação brasileira consolidados pelo QEdU (2007-2015)



Sabendo disso, o BCG GAMMA – braço de *advanced analytics* do BCG – e a Fundação Lemann se uniram para lançar a primeira edição do BCG GAMMA Challenge, um *datathon* que instiga estudantes e jovens profissionais a utilizar *data science* para gerar recomendações de alto impacto para solução de problemas reais. É uma forma de fomentar o desenvolvimento de talentos e o engajamento dos jovens brasileiros em questões sociais, aliando seu potencial intelectual às possibilidades que podem ser exploradas com *analytics*.

“A ideia foi realmente usar um mar de informações que o QEdU compila para tirar insights e ajudar a Fundação Lemann no seu objetivo de melhorar a educação pública no Brasil”.
Henrique Sinatura, Sócio do BCG Brasil

A primeira edição do BCG GAMMA Challenge teve aproximadamente 500 inscritos de todo o Brasil, engajados em resolver o desafio proposto: alavancar *data science* para gerar ideias acionáveis para a Fundação Lemann, utilizando informações do QEdU e de outros dados públicos. Em um curto prazo, os participantes conseguiram extrair *insights* relevantes do material disponibilizado, reforçando hipóteses existentes no setor, trazendo novas perspectivas e propondo ações para alavancar a educação pública brasileira.

“Aqui na Fundação Lemann acreditamos muito no potencial das pessoas em criar soluções inovadoras e no seu efeito multiplicador para transformar o Brasil. Por isso, ver um grupo de jovens extremamente talentosos, acreditando que é possível melhorar a educação do Brasil com o uso de dados e evidências para apoiar a tomada de decisão e o desenho de políticas públicas, é extremamente relevante para nós”. Tamires Vilela, Fundação Lemann

Ideb: contexto e oportunidades em *analytics*

A existência de uma ampla gama de dados é a condição inicial para a aplicação de *analytics* na educação brasileira. Essa oportunidade é potencializada por um segundo fator: a grande dispersão do desempenho (variável de saída) entre as mais de 140 mil escolas públicas no Brasil, conforme medido pelo Índice de Desenvolvimento da Educação Básica (Ideb) e observado na Figura 2. O Ideb classifica as escolas em uma escala de 0 a 10 com base em dois critérios: rendimento escolar (aprovações) e aprendizado em português e matemática.

FIGURA 2 | Há grande variabilidade de performance do Ideb entre as escolas e entre as regiões do Brasil



Fonte: Ideb – INEP; análise de grupo participante do BCG GAMMA Challenge

Grande variância também é observada nas diversas variáveis que podem, por hipótese, esclarecer as diferenças de desempenho. Quanto mais homogêneos e estruturados os dados, mais simples é explicar o comportamento de uma variável de saída, por exemplo, a performance das escolas no Ideb. Porém, quanto mais diversas e não padronizadas são as informações, maior é o valor agregado pela IA para identificar os fatores que melhor explicam o comportamento de cada variável de saída. A Figura 3 indica essa grande variabilidade ao comparar as variáveis educacionais dos 10 melhores e 10 piores municípios em termos de Ideb.

FIGURA 3 | Variáveis de melhor correlação com o Ideb do municípios



% Escolas com todos os anos

Quanto menor a %, maior o Ideb



% Escolas com creche

Quanto menor a %, maior o Ideb



% Escolas com múltiplas dependências

Quanto maior a %, maior o Ideb



Professores por estudante

Quanto maior o #, maior o Ideb



Computadores por estudante

Quanto maior o #, maior o Ideb

Fonte: Ideb – INEP; análise de grupo participante do BCG GAMMA Challenge

Algumas das análises apresentadas pelos participantes durante o BCG GAMMA Challenge confirmam que há *insights* a serem obtidos dos dados de educação que podem influenciar discussões importantes sobre políticas públicas na área. Esses *insights* são um exemplo inicial das possíveis implicações que o uso de *analytics* pode trazer para o setor educacional. A seguir, foram selecionadas três análises realizadas pelos participantes com ideias preliminares de aplicações de *analytics* na educação brasileira.

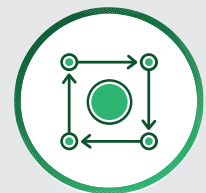
FIGURA 4 | Ideias propostas pelos participantes do BCG GAMMA Challenge



Definição das metas por escola



Progressão das turmas



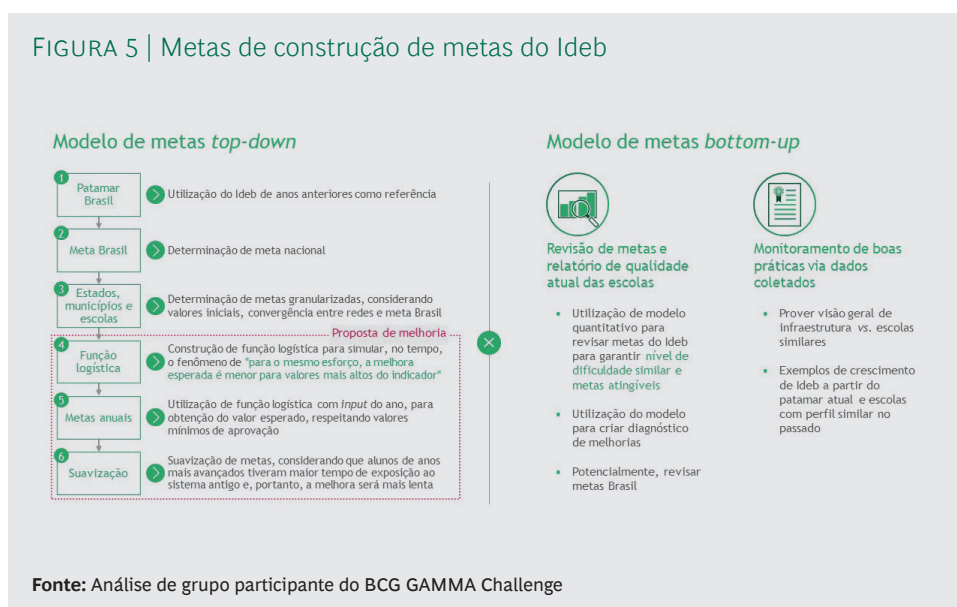
Rotação de professores

Fonte: Análise de grupo participante do BCG GAMMA Challenge

Definição das metas por escola: *machine learning* pode aprimorar projeções

A meta do Ideb, hoje, é projetada de maneira *top-down*: define-se um objetivo nacional e este é cascadeado, de forma granularizada, para os estados, municípios e escolas. Com o uso de *advanced analytics*, é possível projetar metas personalizadas e mais ambiciosas por escola, uma vez que a utilização de modelos preditivos pode indicar a relação do Ideb com outras variáveis, ajudando a definir os fatores de maior impacto e calibrar o potencial de alcance de cada escola. Assim, a meta nacional seria construída de forma *bottom-up* – o que já é uma prática muito utilizada no setor privado.

FIGURA 5 | Metas de construção de metas do Ideb



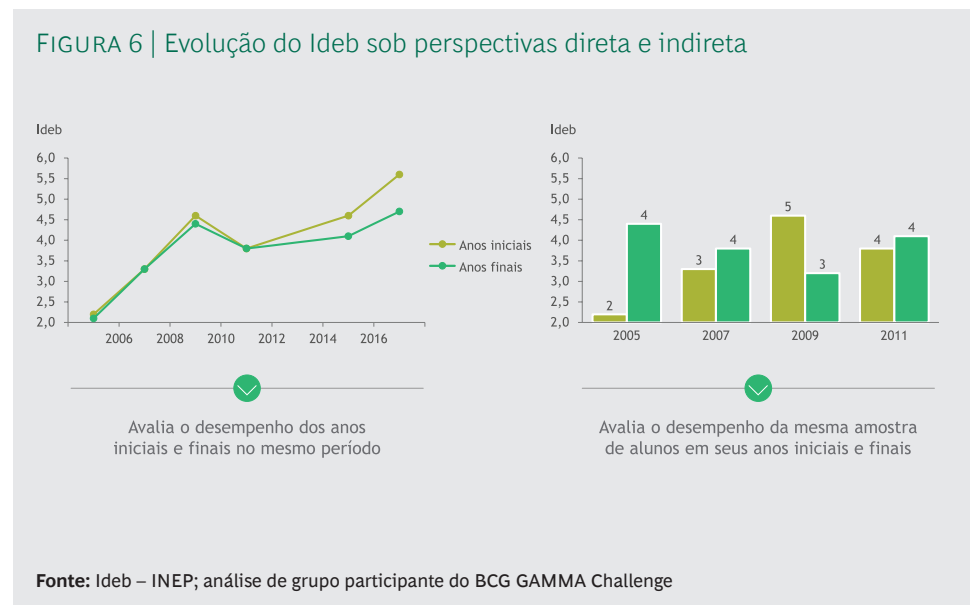
A interpretação do impacto de cada fator na nota do Ideb, como mencionado anteriormente, pode ser feita por métodos como LIME e SHapley Additive exPlanation (SHAP), que ajudam a compreender a motivação do algoritmo por trás de um modelo preditivo criado com *machine learning*. Esse último, em específico, alavanca a teoria dos jogos para entender como fatores concorrentes afetam o resultado do modelo.

Uma das equipes participantes do BCG GAMMA Challenge propôs a utilização de SHAP para estabelecer as variáveis que mais poderiam maximizar a nota do Ideb e aumentar o alinhamento das metas definidas para cada escola. Por exemplo: se a quantidade de professores por aluno e o histórico da escola no Ideb forem variáveis que têm grande impacto na nota do Ideb, é importante considerar que diferentes escolas teriam um potencial distinto de aumentar suas notas de acordo com suas características. Sendo assim, uma meta estabelecida com a ajuda de *machine learning* poderia ser mais calibrada com base na situação atual das escolas vs. potencial de melhoria. Essa análise também poderia ajudar a direcionar os investimentos do município para elementos que têm maior potencial de promover mudança ou para aqueles que possam fazer maior diferença com menor esforço, dependendo do prazo disponível para atingir a meta.

Progressão das turmas: uma visão alternativa à progressão das escolas

Espera-se que o desempenho das escolas melhore continuamente ao longo do tempo. Porém, como avaliar a progressão das turmas? As tendências observadas na progressão do Ideb de uma escola se refletem no nível das turmas?

Um outro grupo participante do BCG GAMMA Challenge argumentou que a tendência de melhora/piora na qualidade da educação ainda não é 100% explícita no Ideb (Figura 6).



O indicador, hoje, pode ser avaliado via correspondência direta, isto é, pela comparação dos anos iniciais e finais no mesmo período. Para que seja possível ter um retrato mais realista da evolução da educação, a sugestão é passar a também analisar o índice via correspondência indireta, ou seja, mesma amostra de alunos no início e final do ensino básico. Se a linha escura é mais elevada que a clara, significa que o desempenho da turma foi pior nos anos iniciais em relação aos anos finais, considerando que os conhecimentos testados são diferentes e apropriados para o momento escolar e que há interferência de outros fatores, como a evasão escolar.

A análise apresentada indica que o modo de avaliar a performance de uma escola ao longo do tempo pode ser revisto. Assim como a evolução das turmas traz *insights* diferentes da evolução da escola, outros modelos de acompanhamento do desempenho podem ser estudados.

Rotação de professores: análise do impacto de uma política pública

A ideia parte da hipótese de que boas escolas têm um corpo docente com melhor formação e maior experiência e que uma escola com pior desempenho se beneficiaria ao receber um professor de uma escola com Ideb maior. Para realizar uma validação inicial dessa hipótese, uma das equipes de cientistas de dados utilizou um modelo de *machine learning* (regressão multivariável) para avaliar a significância do impacto da variável fluxo de professores entre escolas nos resultados do Ideb.

De acordo com a equipe responsável pela análise, foi constatado de forma quantitativa que há um impacto positivo na nota do Ideb quando um professor de uma escola de alto Ideb é realocado para uma escola com Ideb inferior. Entretanto, esse crescimento no Ideb é inferior à realocação no sentido oposto, quando um professor de uma escola com pior desempenho é transferido para uma escola com desempenho melhor. Esse segundo impacto negativo não compensaria o impacto positivo da primeira realocação, no caso desta ser um intercâmbio de professores entre as escolas.

Esse é um exemplo de *insight* poderoso que poderia contribuir para a definição de políticas educacionais e que se beneficiaria de uma análise com maior profundidade e do levantamento de alternativas para minimizar efeitos colaterais – por exemplo, como atenuar o impacto negativo de uma rotação de professores. Também é um exemplo de como *analytics* sempre deve vir acompanhado de um entendimento qualitativo dos assuntos em questão, uma vez que essa conclusão é alvo de debate no meio educacional.

“Compreender o contexto e o problema em questão é tão importante quanto ter domínio das ferramentas para análise de dados. Especialmente em um sistema complexo como o da educação, saber interpretar os dados é fundamental para encontrar as hipóteses corretas e traçar um caminho analítico que nos leve a respostas coerentes – e até novas perguntas”. Henrique Sinatura, Sócio do BCG Brasil

O USO DE *ADVANCED analytics* para analisar dados – que hoje ainda são subutilizados no setor público, seja por falta de infraestrutura (principalmente de TI) ou de *know-how*/cultura – pode gerar *insights* relevantes. Esses *insights* podem trazer à tona novas discussões e formas de gerar valor com políticas públicas, alavancando organizações, explorando dados disponíveis para melhor compreensão do histórico e cenário atual e prevendo resultados futuros. Além disso, torna-se viável a realização de experimentos e sua validação em tempo real, o que possibilita decisões de negócio mais embasadas.

Com apenas duas semanas de trabalho, as equipes que participaram do BCG GAMMA Challenge conseguiram extrair *insights* relevantes sobre o setor de educação e propor melhorias ao atual modelo brasileiro de ensino e de avaliação, mostrando que *analytics* pode ser um forte aliado da gestão pública para direcionar metas, investimentos e a priorização de iniciativas.

As possibilidades de aplicação de *advanced analytics* se estendem a várias outras áreas e setores, como saúde e mobilidade urbana, trazendo soluções personalizadas e diferenciadas que podem representar ganhos significativos e, mais importante, gerar mudança efetiva.

Sobre os Autores

Henrique Sinatura é Sócio no escritório de São Paulo do Boston Consulting Group e líder do BCG Gamma, unidade focada em Advanced Analytics e inteligência artificial. Você pode contatá-lo através do e-mail sinatura.henrique@bcg.com

Tamires Vilela é Gerente de Desenvolvimento Institucional na Fundação Lemann. Você pode contatá-la através do e-mail tamires@fundacaolemann.org.br

Coautores

BCG: Caio Guimarães, Tadeu Rocha, Henrique Coelho, Lissa Wan, Vitorio Canozzi

Fundação Lemann: Aline Okada, Lucas Rocha

QEDu: Michael Mafort

Agradecimentos

O time de autores gostaria de agradecer a Douglas Woods, sócio do BCG, e Michael Mafort, Tech Lead da Fundação Lemann, pelo apoio ao projeto e a esta publicação.

Paras Mais Informações Sobre o Estudo

Caso queira discutir o conteúdo apresentado neste estudo, entre em contato com um dos autores.

Boston Consulting Group (BCG) é uma empresa global de consultoria de gestão e líder em estratégia de negócios. Realiza parcerias com empresas em todos os setores e regiões do mundo com o intuito de identificar as oportunidades que mais geram valor, abordar os desafios mais importantes e transformar o negócio de seus clientes. Nossa abordagem customizada combina amplo entendimento da dinâmica das corporações e de seus mercados com a colaboração de todos os níveis da empresa. Isso garante que nossos clientes atinjam uma vantagem competitiva sustentável, criem organizações mais capazes e garantam resultados duradouros. Fundado em 1963, o BCG é uma empresa privada com 90 escritórios em 50 países. Para obter mais informações, acesse: www.bcg.com.

O time do BCG GAMMA é formado por cientistas de dados e consultores especializados na aplicação de advanced analytics a problemas de negócios complexos, gerando valor expressivo já no curto prazo. O time combina habilidades avançadas em inteligência artificial, estatística, machine learning, otimização e ciência da computação, com profunda expertise nas indústrias dos clientes.

A Fundação Lemann acredita que um Brasil feito por todos e para todos é um Brasil que crê no seu maior potencial: gente. Isso só acontece com educação de qualidade e com o apoio a pessoas que querem resolver os grandes desafios sociais do país. Nós realizamos projetos ao lado de professores, gestores escolares, secretarias de educação e governos por uma aprendizagem de qualidade. Também apoiamos centenas de talentos, lideranças e organizações que trabalham pela transformação social. Tudo para ajudar a construir um país mais justo, inclusivo e avançado. Saiba mais em: fundacaolemann.org.br.

O QEDu, uma iniciativa da Fundação Lemann, nasceu para descomplicar o acesso aos dados educacionais. Acredita que tecnologia e dados são ferramentas essenciais para auxiliar os educadores a tomar melhores decisões pedagógicas e, assim, aumentar o aprendizado dos alunos, garantindo uma educação de qualidade para todos.

© Boston Consulting Group, 2019. Todos os direitos reservados.
07/19

Para acompanhar as publicações mais recentes do BCG sobre este e outros assuntos, acesse www.bcg.com. Também siga BCG in Brazil no Twitter e no Facebook.

