

Explorando os Dados da Educação: ciência de dados para a melhoria do ensino

Exploring Education Data: Data Science for Better Education

INSTITUTO UNIBANCO¹

INTRODUÇÃO

Nosso mundo está ficando cada dia mais conectado e, mesmo que essa transformação não ocorra de maneira democrática, a cada dia são gerados volumes astronômicos de dados sobre tudo que fazemos, gostamos, consumimos e temos interesse. Neste contexto, a ciência de dados tem como objetivo transformar essas informações em ações concretas.

Este especial - que integra o projeto Ciência de Dados na Educação do Instituto Unibanco em parceria com o CRIE (Centro de Referência em Inteligência Empresarial, da COPPE/UFRJ) - é dedicado a explicar como essa ciência pode ser aplicada a partir da exploração dos dados da educação.

Nele, você vai entender como essa abordagem interdisciplinar pode ser usada na educação, requisitos e potenciais benefícios, bem como as questões éticas a serem consideradas na captação e tratamento dessas informações.

Assim, o conteúdo tem como objetivo agregar conhecimento sobre ciência de dados, letramento em dados, ética, dataficação, big data, inteligência artificial, algoritmos, aprendizagem de máquina, entre outros temas, e lhe motivar a ler, trabalhar com, analisar e questionar os dados no seu dia a dia de trabalho

Para uma experiência ainda mais profunda sobre o tema, distribuimos uma série de links para materiais do Observatório de Educação - Ensino Médio e Gestão ao longo do conteúdo. Utilize-os para uma experiência ainda mais rica e completa sobre como explorar os dados da educação e como realizar ações estratégicas através da ciência de dados.

¹ Este artigo faz parte do projeto Ciência de Dados e Inteligência Geográfica, parceria entre o Laboratório de Dados da Gerência de Gestão do Conhecimento (GCC) do Instituto Unibanco e o Centro de Referência em Inteligência Empresarial (CRIE), da Coppe/UFRJ. Publicado originalmente em:

<https://observatoriodeeducacao.institutounibanco.org.br/em-debate/explorando-os-dados-da-educacao>

1. ASPECTOS CONCEITUAIS DA CIÊNCIA DE DADOS E SUAS TECNOLOGIAS

Contribuir para transformar dados em ações. Esse é um dos objetivos da Ciência de Dados. Para isso, são combinados estatística, métodos científicos e diversas técnicas e ferramentas. Essa abordagem interdisciplinar detecta padrões e relações que muitas vezes passariam despercebidos, despertando insights e gerando previsões que subsidiam uma decisão com propósito, embasada em dados e não em achismos.

Na educação, embora sejam extremamente relevantes decisões data-driven, ou seja, orientadas por dados, elas ainda são muito incipientes nas políticas educacionais, quando comparado a áreas como saúde, segurança pública ou mobilidade urbana. Alguns dos motivos para tal cenário podem ser a baixa geração de dados da educação (especialmente os de qualidade e os não capturados por meio de processos tradicionais), a não integração de sistemas, a não padronização de processos e a inexistência de cultura data-driven e letramento digital.

A pandemia de Covid-19 impulsionou a migração para o digital e em muitos casos escancarou as desigualdades digitais que tiveram grande impacto na educação. De algum modo, isso acelerou os [3 V's do Big Data](#) elencados pela empresa americana de consultoria e pesquisa Gartner: o [volume alto de informações](#); a velocidade alta de como os dados são criados, gerados, produzidos e acessados; e a variedade ampla de ativos de informação (dados estruturados, semiestruturados e não estruturados). Também vieram novos desafios, como lidar com o gigantesco volume de dados, produzidos (e desatualizados) em uma velocidade vertiginosa, das mais variadas fontes - incluídos dados sensíveis de estudantes.

Diferentes aspectos da ciência de dados na educação serão abordados ao longo deste especial. Para garantir uma boa experiência de leitura e aprendizado, preparamos, na seção abaixo, um glossário com os principais termos que nortearão o conteúdo. Consulte-o sempre que identificar um termo que não conheça ao longo do restante do texto

2. CONTRIBUIÇÕES DOS DADOS DA EDUCAÇÃO PARA A MELHORIA DO ENSINO

A Ciência de Dados na educação traz diversos efeitos positivos: maior engajamento dos alunos, com ensino customizado; maior possibilidade de traçar estratégias mais efetivas e precisas de [recuperação e aprofundamento](#); realizar análises de cada estudante e ciclos rápidos de melhoria para elaborar essas estratégias.

Para isso, é fundamental desenvolver ações para garantir algumas condições: acessibilidade dos estudantes não somente nas escolas, mas nas suas residências; um ambiente de casa que incentive o ensino e tenha [condições para os estudos](#); e ainda a segurança de renda para que se consiga seguir os estudos. E, por fim, [capacitar o corpo docente](#) em metodologias ativas, tais como aprendizagem por projetos ou entre pares.

Ricardo Henriques, superintendente-executivo do Instituto Unibanco, explica, no vídeo abaixo, algumas dessas questões. Ele participou do [painel "Tecnologias para Redução das Desigualdades: Big Data para o Bem Comum"](#).

A partir de estudos de diversos pesquisadores apresentados em [workshops](#) realizados pela Computing Research Association, Chris Dede, professor da Universidade de Harvard, elencou sete passos para uma educação intensiva em dados. Dede é um nome extremamente relevante na pesquisa de sistemas educacionais que atendam às oportunidades e desafios do século 21. Seus estudos abrangem [desde tecnologias emergentes para aprendizagem](#), políticas que apoiam a transformação educacional, liderança em inovação educacional, ambientes de ensino baseados em mundos virtuais, realidades aumentadas, dentre outros.

7 passos para uma educação data-driven

1. Mobilizar a comunidade para oportunidades com base em novas formas de evidências Chris Dede alerta que a Ciência de Dados, aplicada à educação, não deve ser enquadrada como uma "solução à procura de um problema, mas em vez disso, como uma alavanca para melhorar a tomada de decisão sobre questões perenes no ensino e na aprendizagem". O pesquisador também comenta que nem sempre aqueles que vão produzir os dados serão também os que vão usá-los – e esta distinção nem sempre é clara. É importante pensar em parcerias entre esses dois perfis e alinhar interesses, de modo a garantir o uso eficiente dos dados para otimizar o sucesso dos estudantes. Para isso, uma estratégia seria estabelecer definições comuns sobre conceitos importantes e identificar grupos específicos que irão usar os dados, de modo a desenvolver produtos direcionados a eles.

Segundo o projeto [Data Wise](#) da Harvard Graduate School, existem oito passos no processo de uso eficiente e aprimorado dos dados educacionais, divididos nas macro etapas "preparar", "investigar" e "agir", conforme o modelo abaixo. O projeto foi criado em 1996 para ajudar instituições de ensino a usarem dados de forma colaborativa e com equidade para melhorar a aprendizagem dos estudantes.



Processo de melhoria de dados. Fonte: Data Wise, Harvard Graduate School.

Ao final do passo 8, o processo continua em uma nova rodada de investigação, em espiral. A etapa "preparar" foca em aprender a trabalhar colaborativamente; a etapa "investigar" em aprender a entender o que os estudantes e os professores sabem e o que pode ser feito em conjunto para melhorar o ensino e aprendizagem; e a etapa "agir" foca em agir corretamente e de maneira responsável. O projeto Data Wise recomenda que se incorpore os hábitos mentais ACE (Ação, Colaboração e Evidências) em cada processo. Em outras palavras, trata-se de uma mudança comportamental que envolve o comprometimento coletivo para ação, avaliação e ajustes; colaboração intencional; e foco em evidências.

2. Disseminar a tomada de decisão baseada em evidências em todo o sistema: o uso de dados na educação traz uma série de oportunidades que vão subsidiar a tomada de decisão, seja em pequena escala, como por exemplo feedback em tempo real na sala de aula, ou em grande escala, como intervenções para reduzir custos. Para Dede, uma maneira de disseminar essa abordagem é focar primeiro na pequena porcentagem do corpo docente que está disposto a usar técnicas baseadas em dados e evidências. Então, gradualmente, ficará evidente para os outros professores que os métodos e resultados podem ser replicados por eles. Outra forma de

aumentar a aceitação é passar a mensagem de que a educação baseada em dados e evidências beneficia tanto alunos quanto professores, com melhoria contínua na gestão e no ensino, muitas vezes proporcionando diminuição da carga horária dos professores. A disseminação precisa ser tanto de cima para baixo (com novas tecnologias produzidas e implementadas) quanto de baixo para cima (com professores relatando uma forte necessidade de um novo sistema e de ação). Essa última é uma forma de garantir que as ferramentas produzidas são realmente as com maior demanda. Por fim, é importante ter um conjunto comum e padronizado de avaliações para comparar as experiências, de modo a ter conclusões mais consistentes sobre o uso intensivo de dados em educação.

3. Desenvolver novas formas de avaliação educacional: novas maneiras de medir a aprendizagem podem fornecer novas formas de evidências para a tomada de decisão para estudantes, professores e outras partes interessadas. Valerie Shute, professora de educação da Florida State University, nos Estados Unidos, em relatório editado por Dede, descreve “a coleta contínua de dados como alunos interagindo com ambientes digitais tanto dentro quanto, mais importante, fora da escola. Quando os vários fluxos de dados se aglutinam, as informações acumuladas podem potencialmente fornecer evidências cada vez mais confiáveis e válidas sobre o que os alunos sabem e podem fazer em vários contextos. Isto envolve avaliações de alta qualidade, contínuas e discretas incorporadas em vários ambientes ricos em tecnologia (TREs na sigla em inglês) que podem ser agregados para informar os níveis de competência em evolução de um aluno (em diversas granularidades de dados) e também agregados entre os alunos para informar as decisões de nível superior (por exemplo, do aluno para a classe, para a escola, para o distrito, para o estado, para o país).”

4. Recontextualizar processos de geração, coleta, armazenamento e representação de dados: dados de nível micro (por exemplo, os comportamentos de cada aluno segundo a segundo à medida que aprendem), dados de nível meso (por exemplo, os padrões de ensino dos professores) e dados de nível macro (por exemplo, resultados dos alunos agregados para fins de prestação de contas) são todas entradas importantes para uma infraestrutura de ferramentas e repositórios para compartilhamento e análise de dados abertos. Em artigo, disponível em relatório editado por Dede, o professor da Escola de Pós-Graduação em Educação de Harvard Andrew Ho argumenta que um aspecto importante disso é a “criação de dados”, porque concentra os analistas no processo que gera os dados. Dessa perspectiva, o surgimento do Big Data é o resultado de novos contextos que criam dados, não de novos métodos que extraem dados de contextos existentes. Uma abordagem adicional para determinar quais dados gerar é o Design Centrado em Evidências (ECD, sigla em inglês). Em artigo, Eric Klopfer, professor e diretor do Programa de Formação de Professores Scheller e do Arcade no MIT, aponta como o ECD define quatro modelos relevantes: (1) o modelo do estudante (o que sabe ou pode fazer); (2) o modelo de evidência (o que um estudante pode demonstrar e o que podemos coletar para mostrar o que ele sabe); (3) o modelo de tarefa (a experiência a partir da qual podemos coletar dados); e (4) o modelo de apresentação (como isso realmente aparece para o aluno).

5. Desenvolver novos tipos de métodos analíticos: em artigo disponível em relatório editado por Dedes, Piotr Mitros, especialista em educação que propôs o projeto MITx ao MIT, que levou à criação da plataforma de ensino a distância edX, escreve: “Integrar diferentes formas de dados – desde avaliação por pares à participação em fóruns sociais – resulta em um nível de diversidade dos dados sem precedentes. Isso sugere uma mudança cada vez maior das estatísticas tradicionais para o aprendizado de máquina exige técnicas muito diferentes daquelas desenvolvidas na psicométrica tradicional.” Avanços em métodos analíticos são claramente um avanço necessário para a ciência de dados na educação.

6. Construir capacidade humana para fazer ciência de dados e usar seus produtos: atualmente existem poucos programas de educação em ciência de dados. Fazer a junção de pesquisa educacional com treinamento em ciência de dados ou fornecer uma “trilha” educacional para cientistas de dados poderia oferecer oportunidades interdisciplinares. Dede ainda acrescenta que a ética deve ser incluída em cada etapa do treinamento de ciência de dados para reduzir o dano emocional não intencional, que pode resultar de várias análises.

7. Avançar nas questões éticas, legais e de segurança: é preciso estar atento ao uso ético dos dados desde o início do processo. Isso inclui não apenas questões como privacidade e sigilo, mas também passa por temas como estigma e exclusão. Artigo de Patricia Hammer, do escritório de advocacia PK Legal, no Paquistão, disponível em relatório editado por Dede, indica que cada nova tecnologia que um pesquisador queira usar apresentará uma combinação única de riscos, muitos dos quais podem ser protegidos usando tecnologias disponíveis e políticas de informação adequadas. De maneira geral, a privacidade pode ser protegida adequadamente por meio de servidores e dados criptografados, dados anônimos, acesso controlado aos dados, e aplicação de políticas de privacidade para proteger o acesso não autorizado é excedido aos dados.”

3. APLICAÇÕES PRÁTICAS DA CIÊNCIA DE DADOS NA EDUCAÇÃO

A partir dos conceitos já apresentados, fica claro perceber o quanto as estratégias de Data Science podem ser valiosas para a educação. Considerando os pontos elencados anteriormente, listamos a seguir possíveis aplicações práticas da Ciência de Dados na educação

1. Plataformas adaptativas e ensino customizado: a Ciência de Dados pode ser uma grande aliada dos estudantes tímidos que ficam com vergonha de fazer perguntas ou daqueles entediados com a matéria que já dominam. Os padrões detectados em plataformas para pesquisa, exercícios e testes permitem identificar dificuldades, interesses e habilidades individuais. Com isso, é possível customizar o processo de aprendizagem, tornando-o mais atrativo e eficaz para cada perfil. Cada indivíduo pode, ainda, estudar segundo seu próprio ritmo. Os dados coletados nessas plataformas podem também ser cruzados com dados de redes sociais, a fim de criar ambientes que sirvam para aprender melhor.

A plataforma de gestão do aprendizado Amadeus LMS - do inglês, 'Learning Manager System' ou Sistema de Gestão do Aprendizado, por exemplo, permite um melhor acompanhamento do

desempenho escolar pelos professores, alunos e equipes pedagógicas, por meio de análise e relatórios de avaliação. Ela foi desenvolvida baseado no conceito de Ensino híbrido, ou blended learning, que abordamos no Especial disponível neste [link](#), por uma equipe de pesquisadores e colaboradores brasileiros, dentre eles João Sedraz (Universidade Federal do Vale do São Francisco – UNIVASF) e Fernando da Fonseca de Souza (Universidade Federal de Pernambuco).

O software, de código aberto, já foi implementado em algumas instituições de educação básica, auxiliando no ensino a distância durante a pandemia. Uma delas é o Centro Educacional de Sobradinho (CES), na Bahia, que usa o software na Plataforma Tamoquim, ambiente virtual para realização de aulas e atividades. Professores podem acompanhar, de maneira visual, os alunos que mais interagem, os dias com mais usuários, os recursos mais acessados e os assuntos mais populares. Gestores podem acompanhar, ainda, professores e usuários ativos e dados de avaliação. Alunos podem ter acesso a indicadores individuais e comparados à turma, como tarefas realizadas pontualmente.

2. Desenho de cursos e currículos: a partir da análise de dados obtidos por meio de aplicativos ou ambientes de curso online, por exemplo, é possível identificar conhecimento tácito - aquele que é difícil de ser explicitado, mas que pode ser captado, de algum modo, por tecnologias educacionais. Esse conhecimento pode subsidiar a construção de cursos e currículos. Esses dados podem, ainda, ser cruzados com os tipos de tarefas propostas e de interação dos alunos.

3. Seleção/casamento (matching): a Ciência de Dados pode proporcionar um "Tinder da educação", dando match entre discentes e docentes com cursos e instituições. Isso traz mais possibilidades de satisfação e realização pessoal e facilita o casamento entre pessoas e instituições.

4. Modelos preditivos: com ferramentas de analytics, é possível prever se o perfil de um discente vai se encaixar com o de um curso e instituição ou ainda se um novo currículo vai funcionar com um curso específico. Uma outra possibilidade é identificar indivíduos com alta probabilidade de abandono e, assim, atuar de modo que isso seja evitado.

O Instituto Unibanco vem pesquisando, em parceria com o professor Pedro Costa Ferreira (FGV), modelos de previsão que usam machine learning para alertar quando um estudante está correndo risco de abandonar a escola. Com isso, educadores, secretarias estaduais, coordenadores pedagógicos e diretores poderiam monitorar esses sistemas e desenvolver estratégias para que os alunos se mantenham na escola. Os dados só seriam acessados por pessoas autorizadas, preservando assim a privacidade e levando em conta princípios da LGPD – finalidade, adequação, livre acesso, qualidade dos dados, transparência, prevenção, não discriminação, responsabilização.

O primeiro passo foi identificar fontes externas para coleta de dados relevantes para os modelos preditivos, como Censo Escolar e SAEB. Essas bases oferecem informações sobre infraestrutura da escola, condições de acesso e algumas informações dos alunos, por exemplo. Após a etapa de

limpeza e tratamento dos dados, foram treinados e testados, em escolas do Espírito Santo e Piauí, diversos modelos de regressão logística (logit) e machine learning, tais como ZeroR, One, K Nearest Neighbor, Naive Bayes, Decision Tree, Logit, Bootstrap, Aggregating, Random Forest, AdaBoostM1, Support Vector e Machine.

Em seguida, de modo a avaliar os resultados dos algoritmos foi desenvolvida uma matriz composta por Verdadeiro Positivo (a resposta foi que o aluno era evadido e o aluno realmente era evadido), Verdadeiro Negativo (a resposta do classificador foi que o aluno não evadiu e o aluno realmente não evadiu), Falso Negativo (a resposta do classificador foi que o aluno evadiu, mas o aluno não evadiu) e Falso Positivo (a resposta do classificador foi que o aluno não é evadido, mas o aluno é evadido). A etapa subsequente foi avaliar as métricas de acurácia (representa o percentual das previsões que foram corretamente feitas), precisão (percentual das previsões positivas que estavam de fato corretas, ou seja, de todos os alunos classificados como positivos, qual percentual realmente é positivo?), recall (percentual das observações positivas que foram de fato previstas, ou seja, de todos os alunos que realmente são positivos, qual percentual é identificado corretamente pelo modelo?) e F-score (combina precisão e recall de modo a trazer um número único que indique a qualidade geral do modelo).

No Espírito Santo, do total de abandonos que de fato ocorreram, 72,5% haviam sido corretamente previstos. Ao todo, 73,8% das previsões considerando abandono e não abandono foram acertadas. Entretanto, do total de abandonos previstos pelo sistema, apenas 12,1% de fato aconteceram, ou seja, foram previstos mais abandonos do que os que de fato foram efetivados. O mesmo padrão foi encontrado nos resultados do Piauí, onde 74,6% dos abandonos observados foram corretamente previstos, 67,8% das previsões foram acertadas e 24,9% dos abandonos previstos de fato aconteceram. Assim, os modelos ainda não estão preparados para fazer previsões precisas de abandono. Isso ocorre porque faltam informações relevantes para que eles sejam mais robustos. Os resultados da aplicação de Inteligência Artificial tornam-se mais precisos com mais dados. Sem os dados necessários, nenhum tipo de algoritmo, não importa quão sofisticado, pode funcionar corretamente. Um passo importante agora é sensibilizar para a importância de mais dados (como por exemplo informações socioeconômicas, notas dos alunos e frequência escolar) para o desenvolvimento de sistemas mais detalhados.

5. Otimização: a Ciência de Dados pode contribuir para otimizar a infraestrutura instalada, considerando os desafios atuais. Além disso, pode ajudar a solucionar problemas logísticos relacionados à roteirização (como visita escolar, entrega de material impresso durante a pandemia, etc.). Existem softwares de gestão escolar, como o brasileiro i-Educar, que auxiliam na gestão de processos escolares, matrículas e dados dos alunos. Usado em mais de 100 municípios do Brasil, o i-Educar fornece dados em tempo real sobre servidores – o que permite realizar um gerenciamento de alocações, afastamentos, substituições, faltas e atrasos; sobre transporte escolar – possibilitando realizar o controle de veículos, motoristas, rotas, itinerários e usuários do transporte; sobre horários - ajudando na análise das demandas e disponibilidades de profissionais na rede de ensino em cada período letivo; sobre Educacenso/INEP - o que dá controle de todas as informações para o Censo Escolar com recursos para importação e

exportação automatizada dos dados. O sistema, desenvolvido inicialmente pela Prefeitura de Itajaí-SC, oferece ainda funcionalidades como matrículas, emissão de certificados e diplomas e módulo de biblioteca que faz a gestão de bibliotecas de cada escola.

4. ASPECTOS ÉTICOS DA CIÊNCIA DE DADOS

Em um mundo que avança rapidamente para a digitalização e coletas de dados inteligentes, a proteção à privacidade coloca em pauta restrições ao uso de dados como nunca ocorreu antes. Além disso, padrões mentais inconscientes e discriminatórios estão sendo reproduzidos em escolhas automatizadas de inteligência artificial, dificultando processos democráticos de conscientização e atitudes pró diversidade.

Assim, é fundamental debater como digitalizar a educação sem reproduzir ou aumentar as desigualdades raciais, de gênero e sociais; de que maneira a nova legislação trazida pela LGPD impacta no manejo de dados de programas do meio educacional; quais questões de privacidade estão envolvidas nos acessos às câmeras, telas e microfones quando falamos sobre a proteção de dados de estudantes; e como podemos pensar em uma ciência de dados mais democrática e promotora das igualdades raciais, de gênero e sociais.

A partir das discussões promovidas em diversos encontros, oficinas de criação e na 2ª Jornada IU do Conhecimento sobre Ciência de Dados na Educação, o Instituto Unibanco está elaborando a sua visão sobre LGPD, ética e desigualdades raciais, sociais e de gênero no uso de dados. Esse documento iniciará a construção de um protocolo/agenda de cuidados e restrições para as pesquisas, e demais iniciativas do Instituto Unibanco, que utilizem dados e informações dessa natureza. Confira o que os especialistas têm a dizer sobre o tema:

Combate às desigualdades raciais, sociais e de gênero no uso de dados

Sil Bahia, diretora de programas do Olabi, organização social que trabalha para democratizar a produção de tecnologia, enfatiza que o conceito de diversidade deve ser pensado de forma ampla, incluindo questões de raça, as quais abordamos no conteúdo disponível neste [link](#), e gênero, mas também sociais, geográficas e etárias. Por isso, é importante devolver humanidade às tecnologias, considerando que elas são feitas por pessoas e não são neutras. “Os projetos de inteligência artificial usam dados de um passado opressor, exprimindo séculos de escravidão e colonização e olhares eurocêtricos”, afirmou.

Para ilustrar o viés de algoritmo, citou algumas pesquisas, como a do Instituto de Tecnologia da Georgia, que mostrou que carros autônomos teriam 5% mais chances de atropelar uma pessoa negra. Outro estudo, da Rede de Observatório de Segurança, relatou que 90% das pessoas que são presas por reconhecimento facial são negras.

Diante disso, o Olabi lançou o PretaLab, plataforma que conecta mulheres negras a empresas de tecnologias no país, que é formada predominantemente por homens (68,3%) e pessoas brancas

(58,3%). O PretaLab dialogou com quase 600 mulheres negras para pesquisa e levantamento de dados. Sil Bahia, que coordena o projeto, salientou a importância do levantamento de dados para reflexão e tomada de decisão.

Saiba mais sobre o viés de algoritmo assistindo à [TED Talk de de Joy Buolamwini](#), cientista da computação e ativista digital ganense-americana que trabalha no MIT Media Lab. Joy criou a Algorithmic Justice League, organização que busca combater o preconceito no algoritmo. Sua pesquisa é contada no documentário Coded Bias, disponível na Netflix.

Proteção de dados na educação

Fabro Steibel, diretor executivo do Instituto de Tecnologia e Sociedade (ITS), salientou a necessidade de uma agenda propositiva e salientou que existem dois países, dois “Brasis” que precisam ser igualmente considerados: um Brasil super conectado e um Brasil não conectado. Um tem potencial para educação online, o outro, sofre dificuldades.

Ele ressaltou que estamos naturalizando ações em plataformas de aula ou sessões de terapia online, por exemplo, sem enxergar possíveis problemas futuros. Também chamou atenção para a governança de dados pessoais e algoritmos, através de uma análise pessoal e humanizada sobre cálculos e criação de algoritmos. Pensando em algo que funciona, mas que seja ético, correto e que evite decisões ruins.

LGPD e uso ético dos dados na educação

Vanessa Pirró, da KLA Advogados, apresentou os aspectos mais jurídicos da LGPD, a primeira lei sobre proteção de dados pessoais no Brasil, destacando os seguintes princípios:

Finalidade, que seria avisar ao usuário o porquê que estou usando esse dado;

Necessidade, referente a utilizar os dados que você realmente precisa para uma finalidade específica, ao invés de ter uma avalanche de dados sem utilizá-los;

Transparência, que implica em ser claro e informar ao titular dos dados como e para que eles estão sendo utilizados.

A LGPD traz, ainda, um conceito chamado Privacy by Design, que determina que quando a empresa for pensar na criação de um serviço, produto ou atividade, a privacidade deve existir desde a sua concepção.

Jane Reolo, analista sênior da Gerência de Desenvolvimento de Soluções do Instituto Unibanco, trouxe a temática “A tênue linha entre transparência e exposição – o uso ético dos dados na educação.” Para ela, é fundamental um adensamento reflexivo da LGPD para não apenas reproduzir a dimensão analógica, culturalmente inserida nas escolas. Além disso, a cultura

brasileira se baseia em um controle que acontece pela exposição e pela punição pública e, por isso, temos algo como a LGPD é desafiador. É preciso estabelecer um raciocínio aberto e colaborativo, trazendo alunos para o centro, junto aos gestores de todas as instâncias. Jane trouxe, ainda, quatro dimensões para aprofundar a reflexão sobre a LGPD:

Ética - referente a um conjunto de regras facultativas que tomamos para avaliar o que fazemos e o que dizemos em função do modo de existência que isso implica;

Moral - apresenta-se como um conjunto de regras universais e coercitivas, que consiste em julgar ações e intenções referidas a valores transcendentais (certo e errado; verdadeiro e falso);

Saber - regras codificadas nas quais são estabelecidas relações entre uma forma e outra de saber;

Poder - regras coercitivas nas quais são estabelecidas relações de força com outras forças.

Como garantir o uso ético dos dados da educação

As restrições presenciais impostas pela pandemia impuseram a migração das aulas para o ambiente digital, possibilitando uma maior digitalização de dados dos estudantes, despertando reflexões éticas. Essa virtualização trouxe enormes desafios para a rede, mas também representa uma grande oportunidade de análise para fins de planejamento, personalização ou avaliação escolar.

O Open Data Institute (ODI) define ética de dados como "um ramo da ética que avalia as práticas de dados com potencial de causar um impacto adverso nas pessoas e na sociedade – na coleta, no compartilhamento e no uso de dados". Isso significa adotar, portanto, uma conduta ética no modo como os dados são coletados, usados e compartilhados.

Para ajudar a identificar e gerenciar essas questões éticas, o ODI desenvolveu o Canvas Ética de Dados (Data Ethics Canvas). Trata-se de uma ferramenta com perguntas norteadoras, em 15 blocos, abrangendo quatro áreas: dados, impacto, engajamento e processos (confira na figura abaixo).



Canvas de ética de dados. Fonte: Open Data Institute (ODI)

O canvas foi desenvolvido de modo a se adequar a qualquer contexto, independentemente do tamanho ou escopo do projeto. É baseado no Ethics Canvas, um framework desenvolvido pela ADAPT - Centro de Tecnologia de Conteúdo Digital, que por sua vez teve como base o Canvas de Modelo de Negócios de Alex Osterwalder. Assim, pode ser útil também no contexto educacional.

Confira a seguir alguns questionamentos, que podem ser úteis para promover reflexões sobre a abordagem ética, em relação aos dados dos estudantes.

Em relação aos dados:

1. O que sabemos?
2. Quais são as nossas obrigações (há legislações específicas, etc.)?
3. Quais são os nossos direitos?
4. Que limitações os dados têm?

Em relação ao impacto:

1. O que estamos tentando alcançar?
2. Quais serão os impactos positivos?
3. Quais podem ser os impactos negativos?
4. Como podemos minimizar os impactos negativos?

Em relação ao engajamento:

1. Como interagimos com as pessoas?
2. O nosso propósito está comunicado de modo claro?
3. Estamos sendo abertos e transparentes?
4. Devemos compartilhar nossos dados e resultados?

Em relação aos processos:

1. Que sistemas/treinamentos são necessários?
2. O quão frequente e como fazemos revisões?
3. Quais são as nossas ações?

Segundo o ODI, é possível começar em qualquer lugar do canvas e preencher as seções em qualquer ordem - seja na versão virtual ou na impressa, com post-its. Para cada seção, reserve um tempo para analisar as respostas. Ao final, elabore um plano de ação, com prazo e responsáveis. A ideia é fazer as reflexões propostas pelo canvas toda vez que o contexto mudar.

Sobre o ODI

O Open Data Institute, com sede em Londres, tem a missão de conectar, inspirar e preparar pessoas para inovar usando dados digitais. Foi fundado por Tim Berners-Lee, considerado o criador da WWW, e conta com nodes em todo o mundo. O do Rio de Janeiro é hospedado no CRIE.

Quer saber mais? Acesse o canvas na íntegra em inglês, através deste [link](#).

5. APROFUNDE SEUS CONHECIMENTOS SOBRE IA, BIG DATA E OUTRAS TECNOLOGIAS DA CIÊNCIA DE DADOS

"Dados têm sido chamados de novo petróleo, mas assim como o petróleo, esses ativos têm que passar por pessoas e refinamento para a produção de valor. Isso é letramento em dados", afirmou Jordan Morrow, diretor em Letramento em Dados da plataforma Qlik. Na TED Talk disponível através deste link [Abre em uma nova guia](#), Jordan defende que o letramento em dados abre nossos olhos e nos permite ter sucesso na revolução digital. Ele lista dois "Cs" do letramento em dados: curiosidade e criatividade. Para ele, o computador mais poderoso que existe está nas nossas mentes. Por isso, o elemento humano não deve ser separado dos dados.

Ao longo deste especial, apresentamos várias definições, conceitos e razões para utilizar data science como meio de explorar os dados da educação e extrair ações concretas por meio deles. Se você quer adquirir um conhecimento ainda mais profundo sobre o tema para aplicá-lo em seu trabalho, deixamos abaixo uma série de referências importantes que estão disponíveis aqui no Observatório de Educação. Confira:

Algorithms of Oppression: How Search Engines Reinforce Racism - de Safiya Umoja Noble

A autora apresenta a discriminação de dados como problema social grave da sociedade atual e explica como algoritmos de buscadores como o Google reforçam o racismo e os estereótipos negativos associados aos negros.

[Acesse aqui](#)

Como as democracias morrem - de Steven Levitsky e Daniel Ziblatt

Os autores exploram as bases da democracia e a história de diversos governos autoritários para criar a compreensão de como o enfraquecimento gradual de instituições como o judiciário e a imprensa provocam a ruptura de sistemas atuais.

[Acesse aqui](#)

The Legacy of Inbloom - de Monica Bulger, Patrick McCormick, Mikaela Pitcan

Como uma iniciativa coleta de dados na educação financiada em 100 milhões de dólares pela Fundação Bill & Melinda Gates deu errado e trouxe muito aprendizado

[Acesse aqui](#)

Estudo LGPD & Educação

Publicação lançada pelo CIEB e UNESCO fornece orientações didáticas e aprofundadas sobre a nova legislação e suas exigências

[Acesse aqui](#)

Linha do tempo do racismo algorítmico - por Tarcizio Silva

A timeline é um resultado secundário da pesquisa de doutorado Dados, Algoritmos e Racialização em Plataformas Digitais. Desenvolvido no PCHS-UFABC, o projeto estuda as cadeias produtivas da plataformização digital (mídias sociais, aplicativos, inteligência artificial) e seus vieses e impactos raciais. Os casos, reportagens e reações ao racismo algorítmico podem ser visualizados abaixo e são dados para artigos, conferências, tese e livro em desenvolvimento.

[Acesse aqui](#)

Trabalhos de Cris Dede para implantar a ciência de dados na educação

Conheça mais sobre o trabalho desse pesquisador que definiu os passos para a construção de uma educação Data Driven.

Data-Intensive Research in Education: Current Work and Next Steps - [Acesse aqui](#)

Next steps for “Big Data” in education - [Acesse aqui](#)

Mais informações - [Acesse aqui](#)

Um modelo heurístico para a programação da produção em fundições com utilização da Lógica Fuzzy - Raul Landmann

Descreve a concepção, desenvolvimento e aplicação de uma metodologia heurística para a programação e o controle da produção.

[Acesse aqui](#)