

Sociedade Big Data

SEM CAPACIDADE ANALÍTICA SERÁ DIFÍCIL USUFRUIR DA REVOLUÇÃO PROVOCADA PELA PROLIFERAÇÃO DE DADOS

M

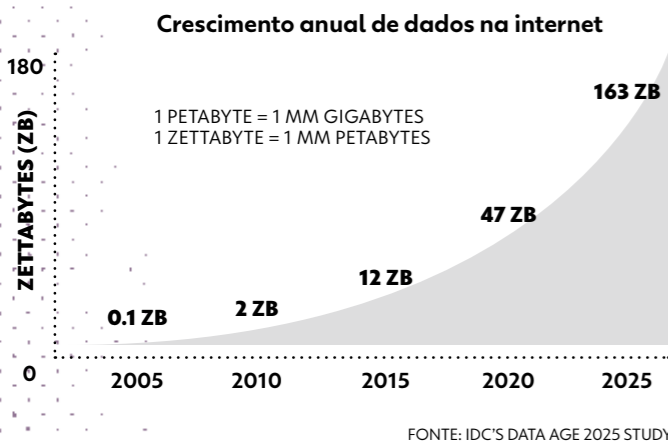
uito se fala em tecnologia exponencial, ou seja, o crescimento não linear de recursos tecnológicos que a sociedade tem à sua disposição. Podemos perceber isso através das ferramentas que estão no nosso dia a dia, como Wi-Fi, smartphones, televisores conectados à internet, relógios inteligentes, entre tantas outras. Mas, enquanto essa tecnologia toda entra em nossas vidas, outra revolução silenciosa ocorre simultaneamente, a Revolução dos Dados.

Todos os recursos tecnológicos que surgiram ao nosso redor trouxeram junto softwares nos quais circulam muitas informações: aplicativos que contam o número de passos, que monitoram a frequência cardíaca, o status de nossos relacionamentos, as fotos do último show a que assistimos, os hábitos de nossos amigos próximos e até daqueles distantes, sem falar em todos os registros de dados em nosso ambiente de trabalho, que conta com mais sistemas para armazenar e disponibilizar informação. No centro de tudo isso estão os dados.

RICARDO CAPPRA é cientista-chefe da Cappra Data Science. Em seu laboratório são criados métodos que ajudam no desenvolvimento analítico de indivíduos e organizações globais como Ambev, Barcelona F. C., Coca-Cola, World Bank, Gerdau, Globo, Santander, Unilever, UOL e muitas outras.



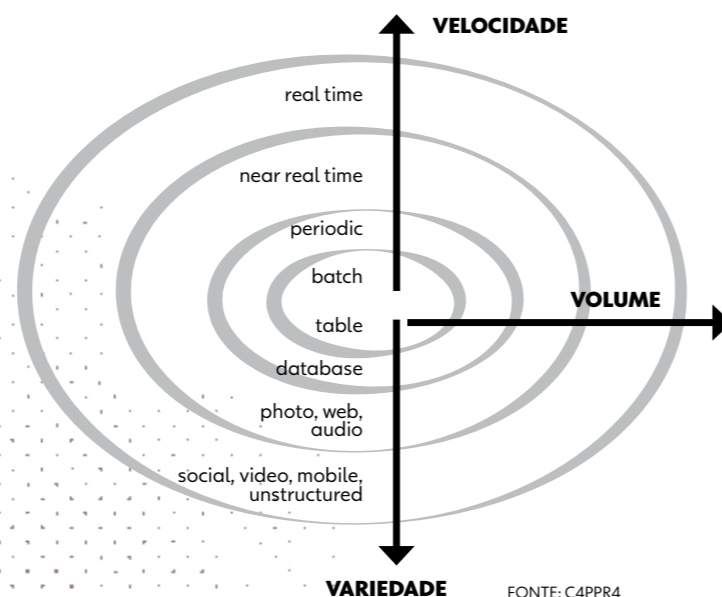
A sociedade rapidamente adotou todos esses recursos tecnológicos, e assim nos tornamos grandes geradores de dados, muitos deles armazenados em bancos de dados de governos (como impostos que pagamos e tantas outras informações que fornecemos), de empresas (registrando todas as nossas atividades e processos) e obviamente na vida pessoal (quando enchemos as redes sociais com novas informações e interações). É muito difícil afirmar o volume de dados que a sociedade está gerando atualmente, isso porque muitos deles estão sob domínio privado, fazendo com que a contagem exata se torne praticamente impossível. Estima-se que, em 2020, a sociedade irá gerar 47 zettabytes (1 zettabyte é o mesmo que 1 milhão de petabytes; 1 petabyte significa 1 milhão de gigabytes), sendo que em 2015 geramos 12 zettabytes, ou seja, vamos quadruplicar em cinco anos. Isso não é nada quando comparado aos próximos cinco anos: em 2025 chegaremos a 163 zettabytes, que representam mais do que dez vezes o número de 2015.



O crescimento contínuo de diferentes tipos de dados, associado ao volume de dados e à velocidade da informação, ganhou o nome de Big Data. Fenômeno que vem causando uma verdadeira transformação em todos os setores da sociedade, sejam eles sociais ou econômicos. O potencial do Big Data está sendo explorado por empresários que criam startups milionárias a partir dos dados que armazenam, pessoas que usam essas informações para influenciar milhões de indivíduos ou, ainda, governos que, a partir dos seus registros, podem tomar decisões que afetam toda a sociedade. É possível afirmar que, hoje, uma das principais fontes do poder é a informação derivada da ciência de dados. Quando entendemos que dados qualificados, organizados e facilmente visualizados podem se transformar em uma fonte inesgotável de informação, fica claro que aqueles que tiverem maior domínio sobre as técnicas de armazenamento, classificação e tratamento dos dados, para transformá-los em informação, terão mais poder.

Para simplificar o entendimento do Big Data, gosto de usar a representação visual abaixo. Ela mostra que o crescimento dos dados em volume, variedade e velocidade causa esse efeito não linear de geração da informação no mundo.

BIG DATA: Crescimento exponencial de dados



Como lidar com tanta informação?

Esse é, sem dúvida, um dos principais desafios atuais. Com tantos dados disponíveis, como definir quais armazenar, quais usar para gerar informação e quais descartar?! Recentemente estudamos em nossos laboratórios o impacto desse consumo de informação não qualificada em nossas vidas e descobrimos uma doença, a intoxicação por consumo excessivo de informação, que ganhou o apelido de *infoxication*.

O dr. David Lewis, Ph.D. em psicologia, documentou no livro *Information overload*, de 1999, um estudo sobre a síndrome da fadiga por consumo de informação (*information fatigue syndrome*), em que verificou o impacto do consumo excessivo de informação na vida das pessoas, comprovando que ele afeta diretamente nosso sistema imunológico por causa do bombardeio a que nosso cérebro é submetido a processar diariamente. Toda essa poluição de dados gera uma verdadeira "paralisia analítica". Ou seja, afeta nossa capacidade de distinguir os fatos que são realmente relevantes para uma análise adequada, prejudicando nossa tomada de decisões. Tudo isso, segundo o trabalho de Lewis, interfere em nosso sono e capacidade de concentração, e os sintomas dessa fadiga podem aparecer em forma de depressão, ansiedade, irritabilidade, estresse e até sintomas físicos como a fibromialgia.

Em resumo, ter muitos dados para tomar decisões nem sempre é o melhor caminho. É preciso transformar a imensa massa de dados ao nosso alcance em poucas análises relevantes, e assim, diante de uma visão mais clara dos cenários disponíveis, escolher as melhores alternativas. Para isso, precisamos desenvolver nossa capacidade analítica, criar técnicas e mecanismos para usufruir do melhor dessa revolução que o efeito Big Data está causando.

TER MUITOS DADOS PARA A TOMADA DE DECISÃO NEM SEMPRE É O MELHOR CAMINHO. É PRECISO TRANSFORMÁ-LOS EM ANÁLISES RELEVANTES

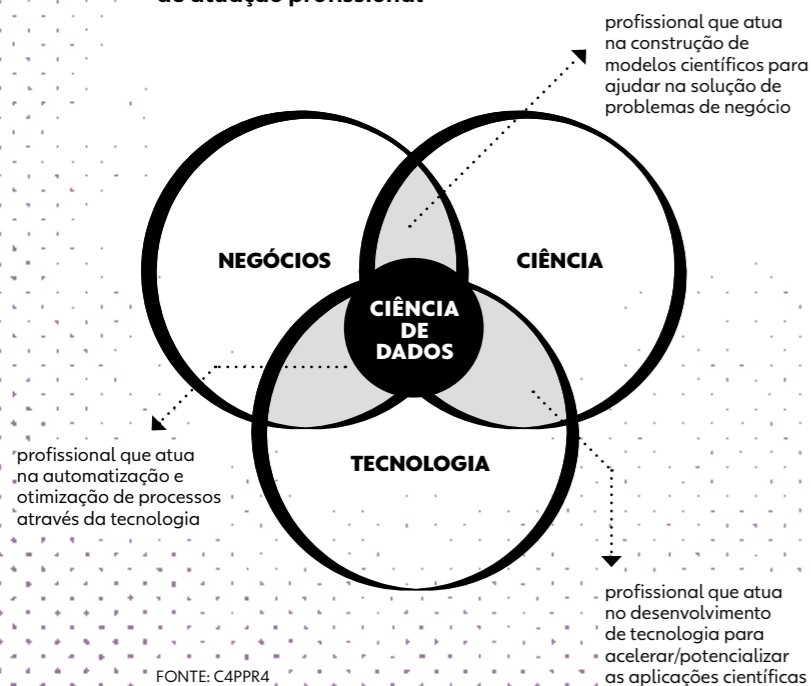
Uma falha sistêmica

Especialmente no Brasil, nossa capacidade de análise foi pouco desenvolvida. Na educação básica, as escolas ensinam matemática e estatística de forma mecânica, baseada na simples repetição, fazendo com que os alunos decorem as fórmulas e não se preocupem com a lógica. Isso significa que a modelagem analítica por trás de uma fórmula não é repassada da forma adequada, prejudicando assim a formação básica nessa área. Em resumo, o sistema de educação básica forma uma população menos analítica, com menor desenvolvimento lógico; uma herança que carregamos para o resto de nossas vidas. Nos países desenvolvidos, as disciplinas que chamamos de exatas são ensinadas de forma diferente, destacando-se os motivos das composições das fórmulas, o que gera maior senso crítico no desenvolvimento do pensamento sistêmico e analítico.

A capacidade analítica menos desenvolvida é, sem dúvida, uma barreira para elevar o nível de pesquisa e desenvolvimento de métodos de análises e, conseqüentemente, o uso de todos esses dados disponíveis, pois trabalhar com dados não deveria ser um privilégio da área de tecnologia. A grande verdade é que as melhores práticas em ciência de dados costumam ser compostas por três pilares de conhecimento: ciências exatas, tecnologia e negócios.

Os profissionais que atuam diretamente no processo de transformação de dados em informação são conhecidos como cientistas de dados, a profissão mais desejada e mais bem remunerada dos EUA, quando comparada com todas as outras áreas do conhecimento, segundo o Glassdoor, um dos mais importantes portais americanos sobre carreiras. Eles costumam atuar aplicando sua expertise entre as grandes áreas de conhecimento, como pode ser visto neste gráfico:

Disciplina Ciência de Dados e vertentes de atuação profissional



ALGORITMOS SÃO REPETIDORES DE REGRAS HUMANAS

Mas a capacidade analítica não deveria estar restrita a um grupo de profissionais. Ela deveria fazer parte de nossa formação básica. Afinal, em um futuro muito próximo, precisaremos inevitavelmente ser mais analíticos. Pense um pouquinho nos últimos aplicativos que você baixou em seu smartphone, seja um app de prática esportiva, um game ou algo para o ajudar no deslocamento de sua casa para o trabalho. Todos eles possuem recursos para você mesmo analisar os dados. Os fornecedores de tecnologia já estão transferindo as análises individuais para os usuários, permitindo que cada um faça seus próprios estudos baseados nos seus dados primários, combinados com dados de terceiros.

Análises, algoritmos e inteligência artificial

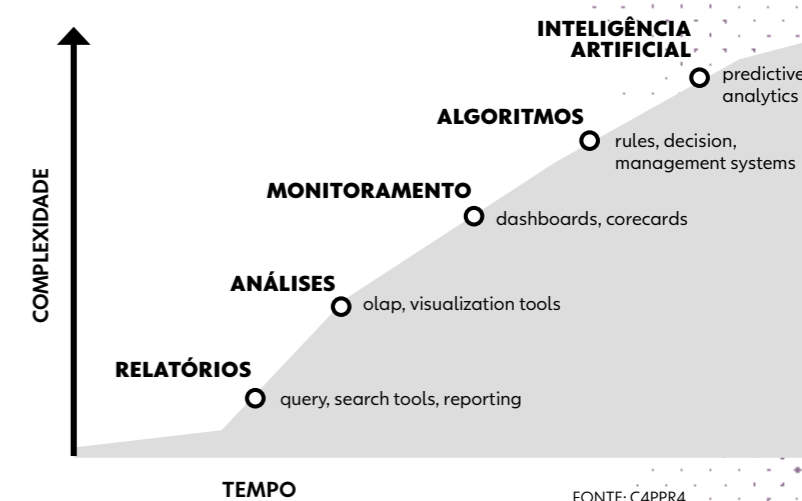
Quando cito o termo “capacidade analítica”, isso não está restrito à análise de dados. Na verdade, o primeiro passo para o desenvolvimento de um algoritmo, por exemplo, não é o código que será criado para o funcionamento dele, e sim as regras que a máquina deverá seguir. Ou seja, a lógica. Um algoritmo nada mais é do que uma série de regras estruturadas

em uma sequência de passos predefinidos. Não são os algoritmos que são inteligentes. A inteligência está nas regras que eles utilizam. O código de computador é apenas um roteiro para que as regras aconteçam. Algoritmos são repetidores de regras humanas. Obviamente, um grande número de dados devidamente estruturados, com sequência bem definida, aliado à capacidade de processamento das máquinas de hoje, cria algo que o humano não tem capacidade de fazer sozinho. O uso da tecnologia é fundamental, mas sem uma lógica bem estruturada, ela não passará de algoritmos (regras), passíveis de falha.

A inteligência artificial, na qual somos instados a prestar atenção em tudo que é jornal, revista, noticiário, filme e séries de TV, nada mais é do que um algoritmo com um processo de aprendizagem constante que, a partir da codificação inicial, armazena o aprendizado em suas próprias bases de dados, para a partir dele tomar “suas próprias decisões” e evoluir com uma capacidade de processamento infinitamente maior que a do ser humano. Coloquei as aspas em “suas próprias decisões” porque precisamos lembrar que as regras iniciais para aprendizagem/execução são inseridas pelos programadores dessa inteligência artificial, ou seja, por seres humanos, criadores das regras. Quanto maior a capacidade analítica desse humano, melhor funcionarão.

A representação gráfica abaixo mostra a evolução que nos trouxe ao ponto em que estamos hoje na construção de algoritmos e da inteligência artificial. O início remonta ao consumo de relatórios, construção de análises, monitoramento de informação... Tudo isso permite armazenar melhores dados, para construir melhores regras, e possibilitar a execução pelas máquinas. Um caminho totalmente baseado na capacidade analítica do ser humano, que cria a lógica de funcionamento das máquinas.

Evolução analítica



PENSAMENTO ANALÍTICO, HABILIDADES ANALÍTICAS E ORIENTAÇÃO POR DADOS SÃO TRÊS FUNDAMENTOS A SEREM DESENVOLVIDOS NOS PRÓXIMOS ANOS

Negócios orientados por dados

Assim como a tecnologia, softwares e dados estão presentes em tudo que está ao nosso redor. E, mais recentemente, também em todas as etapas dos negócios. Empresas usam rastros digitais deixados por seus consumidores (nós!), gerando recomendações, em suas próximas visitas, baseadas em seus hábitos e preferências individuais. Isso faz com que aumente muito a precisão do marketing, por exemplo, que vai entregar mensagens cada vez mais adequadas ao seu público, permitindo que ele mesmo faça a personalização do que é mais relevante. As recomendações mais refinadas serão muito valorizadas pelos consumidores no meio dessa avalanche de dados. O conteúdo mais relevante e personalizado permitirá que façamos uma melhor gestão do nosso tempo e reduzirá o efeito da *infoxication*. Obviamente, precisaremos de regras claras para uso de dados, respeito da privacidade do indivíduo e consenso de quando essa troca valerá ou não a pena. A cessão dos dados será uma opção do usuário, que precisará ter em suas mãos a possibilidade de trocar dados individuais por conveniência e/ou produtos e informações melhores. Esse debate já acontece no mundo todo.

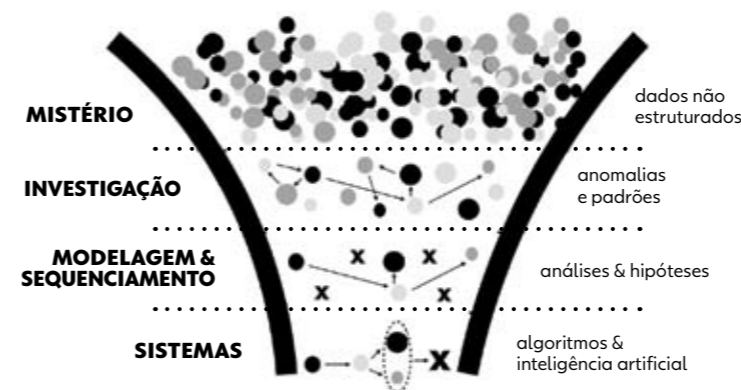
Quando pensamos em processos nas áreas de negócio, são muitos os incrementos de valor que já podemos encontrar no mercado: algoritmos que sugerem melhores rotas de entregas, baseados no histórico e trânsito em tempo real; modelos matemáticos que recomendam o melhor momento de fazer a compra de determinada matéria-prima, baseados em buscas na internet; dados de mercado e monitoramento em tempo real de taxas e câmbios; sistemas de recomendação para recrutamento e seleção de profissionais baseados em seus dez últimos anos de carreira... No dia a dia, isso representa menos decisões baseadas em *feeling* (percepções individuais) e mais decisões *data-driven* (orientadas por dados), permitindo que deixemos de lado as tarefas repetitivas e nos concentremos naquilo que é realmente mais importante para nós.

Nada do que falamos aqui diminui a importância das pessoas envolvidas nesses processos de negócio. O uso de dados não vai concorrer com os bons profissionais que atuam em logística, compras, recursos humanos. Na verdade, os tornará mais produtivos através do uso de ferramentas de automação, para que o esforço braçal/operacional/mecânico seja menor, e para que eles possam usar mais de sua capacidade analítica na criação de regras melhores. Assim, poderão elevar a precisão do processo e gerar valor através da tecnologia, que tem capacidade de digerir muito mais informação, muito mais rápido, com muito menos chance de erro.

O processo de transformação dos dados

Para representar um pouco desse processo de transformação, vou usar um funil. Na parte superior veremos muitos dados, variados, em um banco não estruturado. Essa camada carrega um *Mistério*, pois é muito difícil saber se os cruzamentos desses dados levarão realmente a resultados relevantes. A próxima camada representa o momento de *Investigar*, cruzar todas as variáveis possíveis, aplicar fórmulas e modelos que ajudem a identificar padrões e anomalias (termo usado para identificação de variâncias nos padrões). Na fase seguinte, de *Modelagem & Sequenciamento*, nascem os modelos de análise específicos, onde hipóteses são testadas e as regras que comporão os sistemas inteligentes são organizadas. Só na última camada serão desenvolvidos os sistemas, algoritmos e rotinas com inteligência artificial.

Intelligence funnel: transformação dos dados em inteligência



FONTE: C4PPR4

Não existe atalho, é um processo realmente trabalhoso, que exige tempo e dedicação para ser desenvolvido. Como citei antes, não se trata de um código de computador complexo, e sim de muita investigação, baseada em análises, para concepção de um sistema lógico que precisa ser testado até a exaustão, para então se transformar no código que a máquina executará.

Pensamento analítico, habilidades analíticas e orientação por dados são três fundamentos que precisaremos desenvolver nos próximos anos. A transformação digital já impactou todas as pessoas no mundo e essa próxima revolução silenciosa já está acontecendo. A Revolução Analítica está sendo dominada por aqueles que adotaram um pensamento analítico, que em razão disso estão acelerando o desenvolvimento de suas habilidades analíticas, independentemente das posições que ocupam, e estão criando uma cultura orientada por dados em suas áreas de atuação, através de melhores práticas analíticas.

E você, vai ficar de fora dessa? ●