

# Agora<sup>1</sup>

Em 2009, um novo vírus da gripe foi descoberto. Combinando elementos dos vírus que causavam a gripe aviária e suína, este novo tipo, chamado H1N1, se espalhou rapidamente. Em semanas, órgãos de saúde pública em todo o mundo temeram que uma terrível pandemia estivesse a caminho. Algumas pessoas previram um caso como o da gripe espanhola de 1918, que infectou meio bilhão de pessoas e matou dezenas de milhões. Pior, não havia vacina disponível contra o novo vírus. A única esperança que as autoridades de saúde tinham era desacelerar a propagação. Mas, para isso, elas precisavam saber onde estava a doença.

Nos Estados Unidos, os CDCs (*Centers for Disease Control* - Centros de Controle e Prevenção de Doenças) pediram que os médicos informassem todos os novos casos de gripe. Ainda assim, o cenário obtido da pandemia estava sempre defasado em uma ou duas semanas. As pessoas podiam se sentir mal por dias, mas tinham de esperar para se consultar com um médico. A transferência das informações para as organizações centrais levava tempo, e o CDC só computava os números uma vez por semana. Com a rápida disseminação da doença, o intervalo de duas semanas era uma eternidade. Este atraso impediu que as agências de saúde pública obtivessem o real panorama da pandemia em momentos cruciais.

Poucas semanas antes de o vírus H1N1 aparecer nas manchetes, engenheiros da **Google** publicaram um notável trabalho na revista científica *Nature* que causou comoção entre autoridades de saúde e cientistas da computação, mas que, de resto, foi ignorado. Os autores explicaram como a **Google** pôde “prever” a disseminação da gripe de inverno nos Estados Unidos, não apenas nacionalmente, mas em regiões específicas e até mesmo em estados. A empresa obteve essa previsão ao analisar os termos mais pesquisados na internet. Como o **Google** recebe mais de três bilhões de pesquisas por dia e as salva, a empresa tinha muitos dados com os quais trabalhar.

A **Google** pegou os 50 milhões de termos de busca mais comuns dos americanos e os comparou a lista com os dados do CDC sobre a disseminação da gripe entre 2003 e 2008. A ideia era identificar pessoas infectadas pelo vírus da gripe de acordo com o que pesquisavam na internet. Outros haviam tentado fazer o mesmo com os termos de busca, mas ninguém dispunha de tantos dados, poder de processamento e conhecimento estatístico como a **Google**.

Apesar de os funcionários da **Google** pensarem que as buscas poderiam ser usadas para a obtenção de informações sobre a gripe - ao escrever frases como “remédio para tosse e febre” -, essa não era a questão: eles não sabiam e criaram um sistema sem muita utilidade. Tudo o que ele fazia era procurar correlações entre a frequência de certas buscas e a disseminação da gripe em relação ao tempo e ao espaço. No total, eles processaram impressionantes 450 milhões de modelos matemáticos diferentes a fim de testar os termos de busca, comparando suas previsões com os casos reais de gripe registrados pelo CDC em 2007 e 2008. E tiveram sorte: o programa descobriu uma combinação de 45 termos de busca que, quando usados juntos num modelo matemático, tinham forte correlação entre a previsão e os números oficiais. Como o CDC, eles podiam ver por onde a gripe havia se espalhado, mas, ao contrário do CDC, apontar a disseminação quase em tempo real, e não com uma ou duas semanas de atraso.

Assim, quando a crise do vírus H1N1 ocorreu, em 2009, o sistema da **Google** provou ser um indicador mais útil e pontual que as estatísticas do governo, com seu atraso natural. As autoridades de saúde pública estavam munidas de valiosas informações.

Incrivelmente, o método da **Google** não envolve o uso de amostras de saliva ou o contato com médicos. Ele se baseia em “big data” - a capacidade de uma sociedade de obter informações de maneiras novas a fim de gerar ideias úteis e bens e serviços de valor significativo. Assim, quando a próxima pandemia surgir, o mundo terá um instrumento melhor à disposição para prever e, assim, evitar a disseminação da doença.

A saúde pública é apenas uma área na qual o big data está fazendo a diferença. Setores da economia também estão sendo reformulados. A compra de passagens aéreas é um bom exemplo.

Em 2003, Oren Etzioni precisava viajar de Seattle para Los Angeles, para o casamento de seu irmão caçula. Meses antes da cerimônia, ele entrou na internet e comprou uma passagem aérea, acreditando que, quanto antes o fizesse, mais barato seria o bilhete. Durante o voo, por curiosidade, ele perguntou à pessoa ao seu lado quanto ela pagam pela passagem e quando a comprara. O homem havia pagado consideravelmente menos que Etzioni, mesmo tendo comprado a passagem recentemente. Furioso, Etzioni perguntou aos demais passageiros. A maioria deles havia pagado menos.

---

<sup>1</sup> extraído do Capítulo 1 do livro “Big Data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana”; Viktor Mayer-Schonberger e Kenneth Cukier; 1a. ed. - Rio de Janeiro, Elsevier, 2013.

Para a maioria de nós, a sensação de traição econômica teria se dissipado assim que tivéssemos recolhido as bandejas e voltado as poltronas à posição vertical. Mas Etzioni é um dos principais cientistas da computação dos Estados Unidos. Ele vê o mundo como uma série de problemas de dados - problemas que ele é capaz de resolver. E Etzioni os domina desde que se formou em Harvard, em 1986, ao se tornar mestre em Ciência da Computação.

De sua varanda na *University of Washington*, ele deu início a várias empresas de “big data” antes de o termo se tornar conhecido. Ajudou a criar um dos primeiros mecanismos de busca da Internet, o *MetaCrawler*, lançado em 1994 e comprado pela *InfoSpace*, então uma grande empresa da área. Ele cofundou o *Netbot*, primeiro grande site de comparação de preços, que vendeu para a *Excite*. Sua startup para extrair significado e documentos texto, chamada *ClearForest*, mais tarde foi adquirida pela *Reuters*.

De volta à terra firme, Etzioni estava determinado a descobrir uma maneira de as pessoas saberem se o preço das passagens on-line é ou não um bom negócio. Um assento de avião é uma *commodity*: cada um é basicamente indiferenciável dos demais assentos do mesmo voo. Ainda assim, os preços variam muito, com base em vários fatores, na maior parte conhecidos apenas pelas companhias aéreas.

Etzioni concluiu que não precisava decifrar o enigma ou o motivo para a diferença de preço. Ao contrário, tinha de prever se os preços mostrados aumentariam ou diminuiriam no futuro. Era possível, ainda que não fosse fácil.

Era preciso analisar todos os preços de passagem de uma rota e examinar os preços pagos em relação aos dias anteriores ao voo.

Se o preço médio de uma passagem tendia a diminuir, fazia sentido esperar e comprar a passagem mais tarde. Se o a tendência era aumentar, o sistema recomendaria que se comprasse a passagem imediatamente. Em outras palavras, era preciso uma versão estendida da pesquisa informal que Etzioni fizera a 30 mil pés. Obviamente, tratava-se de outro enorme problema ciência da computação, mas que ele era capaz de resolver. Então Etzioni se pôs a trabalhar.

Com uma amostragem de 12 mil preços obtidos em um website de viagens durante 41 dias, Etzioni criou um modelo de previsão que proporcionou grande economia aos passageiros simulados. O modelo não entendia o porquê, apenas o quê, isto é, ele não conhecia as variáveis por trás da decisão das companhias aéreas, como a quantidade de assentos não vendidos, a temporada ou a mágica possibilidade da redução do preço em uma noite de sábado. O sistema baseava suas previsões no conhecido: probabilidades obtidas a partir de dados de voos.

“Comprar ou não comprar, eis a questão”, pensou Etzioni. Assim, ele chamou o projeto de pesquisa de Hamlet.

O pequeno projeto evoluiu para uma *startup* de capital de risco chamada *Farecast*. Ao prever se e quanto o preço de uma passagem aérea aumentaria ou diminuiria, a *Farecast* possibilitou que os escolhessem quando apertar o botão de compra. O sistema os munuiu de informações às quais nunca tiveram acesso. De acordo com um princípio de transparência, a *Farecast* até mesmo anunciava o grau de precisão de suas previsões e também apresentava essas informações ao público.

Para funcionar, o sistema precisava de muitos dados. Para melhorar seu desempenho, Etzioni conseguiu um dos bancos de dados de reservas da indústria aeroviária. Com essa informação, o sistema podia fazer previsões com base em casa assento, em cada voo, na maioria das rotas da aviação comercial americana ao longo de um ano. A *Farecast* usava agora quase 200 bilhões de registros de preços para fazer previsões. Assim, economizava muito para os usuários.

Com seus cabelos castanhos, sorriso aberto e aparência de querubim, Etzioni não parece a pessoa que negaria à indústria aeroviária milhões de dólares em faturamento potencial. Na verdade, ele tenta fazer mais que isso. Em 2008, ele planejava aplicar o método a outros bens, como quartos de hotel, ingressos de shows e carros usados: qualquer serviço com pouca diferenciação, alta variação de preço e toneladas de dados. Mas antes que pudesse dar continuidade a seus planos, a **Microsoft** bateu à sua porta, comprou a *Farecast* por cerca de US\$110 milhões e a integrou ao sistema buscas **Bing**. Em 2012, o sistema acertava 75% das previsões e os passageiros economizavam, em média, US\$50 por passagem.

A *Farecast* é um ícone de empresas de big data e um exemplo de para onde o mundo está indo. Cinco ou dez anos antes, Etzioni não a poderia ter criado. “Teria sido impossível”, diz ele. A quantidade de poder e processamento e de armazenamento de que ele precisava era cara demais. Mas apesar de as mudanças na tecnologia terem sido um fator crítico que possibilitou o negócio, algo mais importante e sutil também mudou: a mentalidade sobre como os dados poderiam ser usados.

Os dados não eram mais considerados estáticos e banais, cuja utilidade terminava depois que o objetivo da coleta era alcançado, como no pouso do avião (ou, no caso da **Google**, após o processamento da busca). Em vez disso, os dados se tomaram matéria-prima dos negócios, um recurso econômico vital, usado para criar uma nova forma de valor econômico. Na verdade, com a mentalidade certa, os dados podem ser reutilizados para se tomarem fonte de inovação e novos serviços. Eles podem revelar segredos para aqueles com humildade, disposição e instrumentos para ouvir.