

# MAPEANDO O FUTURO COM *ANALYTICS*

POR **MIGUEL VILLELA, AGOSTINHO VILLELA E HUGO TADEU**

Sobrevivência e competitividade. São esses, acima de todos os outros, os objetivos fundamentais de qualquer organização. Até porque, sem eles não é possível pensar nos demais, como o ambiente saudável de trabalho, atendimento ao cliente, produtividade, crescimento e, muito menos, no mais importante de todos – a lucratividade. Sempre foi assim, e assim será.

No cotidiano atual, marcado por rupturas tecnológicas e de mercado cada vez mais frequentes, com ciclos de maturação sempre mais rápidos, qual é a principal ferramenta das empresas para sobreviver e competir? Temos, há bastante tempo, a percepção de que é possível extrair informações significativas a partir de dados armazenados. Assim, com a

verdadeira “selva de dados” em que se transformou o mundo moderno, as organizações se voltaram para o desenvolvimento de técnicas que as ajudem a extrair informações cada vez mais sofisticadas. Tentam, assim, entender da melhor forma possível o comportamento do seu mercado e de como sobreviver e competir nele.

A essas técnicas de análise de dados foi dado o nome de *analytics*. Com o tempo, elas foram se aprimorando de tal forma, que passaram a colaborar também com outros objetivos das organizações. E, hoje, já impactam todo o ambiente organizacional. Praticamente, não existe área do mundo corporativo que não seja, ou não possa vir a ser, beneficiada pelo uso de *analytics*.

Cabe ressaltar que, apesar de ter se tornado de uso comum – assim como outros termos empregados em nosso cotidiano, como *big data*, *machine learning*, *data science*, etc. –, a definição de *analytics* ainda é meio vaga. Segundo a Wikipédia, “é a descoberta, interpretação e comunicação de padrões significativos de dados”. No entanto, a própria enciclopédia livre afirma que ainda não está clara a diferença entre *analytics* e análise de dados.

Independentemente disso, com a evolução do volume de dados disponível hoje, no mundo, a inteligência analítica vem sendo utilizada de forma ampla em todos os níveis das organizações, sobretudo as de maior porte.

**PRINCIPAIS USOS DE ANALYTICS** O leque de aplicações de *analytics* vai desde o processo de tomada de decisão pela alta gerência até a definição de produtos por loja, execução de estratégias de marketing, dimensionamento da equipe de vendas, otimização de estoques e prevenção de fraudes. Vejamos isso mais detalhadamente:

- **Tomada de decisões em gestão** – a utilização de *analytics* na tomada de decisões permite, sobretudo, identificar e executar oportunidades que, a olho nu, aparecem subvalorizadas, reagir de forma rápida a mudanças de mercado e aprender a lidar com bases de dados, em vez de se basear em preconceitos.
- **Otimização, modelagem de preços e promoções** – o preço é extremamente importante no estabelecimento da imagem de uma marca e o que ela representa. Aquela que consegue se conectar com seus consumidores, por meio da mensagem de preço, garante a lealdade do cliente no longo prazo. O preço é também a alavanca com maior impacto imediato na maximização do lucro e, quando gerenciado ao longo do tempo, leva a um saudável crescimento dos lucros. Executivos do varejo entendem, de forma clara, a capacidade de melhoria de margens oferecida por *analytics* de preço. De acordo com o relatório *Pricing 2015*, da RSR Research, os varejistas consideram a melhoria de margens como uma das três maiores oportunidades de contribuição da precificação para a estratégia do negócio. Vêm, em seguida, a criação de promoções mais lucrativas e a sustentação de uma imagem mais competitiva.

- **Dimensionamento e otimização da equipe de vendas** – um exemplo claro de dimensionamento da equipe de vendas é obtido com a metodologia utilizada pela IBM para aprimorar a eficiência e produtividade de sua equipe de vendas, em nível mundial, que utiliza processos de otimização e analítica quantitativa. Essa metodologia é implementada através de três iniciativas abrangentes: programa de crescimento e desempenho (GAP – *Growth and Performance*), programa de otimização territorial (TOP – *Territory Optimization Program*) e otimização de cobertura com lucratividade (COP – *Coverage Optimization with Profitability*). O GAP oferece um conjunto de modelos analíticos para mensurar e otimizar a capacidade de venda e o crescimento com lucratividade. O TOP desenvolve modelos analíticos e métodos para análise e otimização na alocação de clientes a vendedores e outros canais de vendas. Por fim, o COP oferece recomendações adicionais de ajustes na cobertura de vendas, com base em melhorias na estimativa de lucro por cliente.
- **Otimização de estoques e perfil de produtos por loja** – para compreender o real valor de cada produto do portfólio, as companhias de distribuição industrial precisam reconhecer que as associações entre produtos não são simples métricas para o gerenciamento de produtos, ou SKUs (*Stock Keeping Units*). A maioria dessas empresas mensura o desempenho de vendas e margem bruta para cada SKU. Mas essa visão não permite uma noção precisa do valor de cada SKU para todo o portfólio. Embora o valor no nível individual de produto seja fundamental, quando as companhias fazem um exercício de racionalização, para determinar os SKUs a serem mantidos, excluídos ou reprecificados, normalmente visualizam cada produto de forma estática, sem levar em conta sua associação com os demais. A informação adicional dessas associações de produtos permite a construção de *mixes* mais efetivos, precificação com maior precisão e gerenciamento de portfólios mais efetivos, através de melhor compreensão da atração e rejeição entre produtos. Algoritmos podem ser utilizados para identificar e mensurar a probabilidade de determinado conjunto (um ou mais produtos) ser encontrado num carrinho, contendo outro conjunto. Os resultados desses algoritmos são robustos e podem ser altamente reveladores.

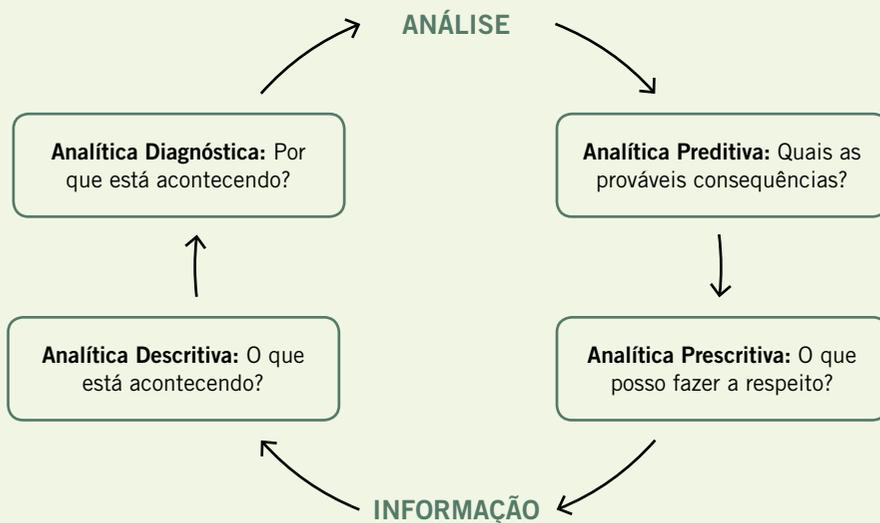
## COM A EVOLUÇÃO DO VOLUME DE DADOS DISPONÍVEL HOJE, NO MUNDO, A INTELIGÊNCIA ANALÍTICA VEM SENDO UTILIZADA DE FORMA AMPLA EM TODOS OS NÍVEIS DAS ORGANIZAÇÕES

No entanto, é importante ressaltar que serão sempre algoritmos extremamente pesados, já que, mesmo para distribuidores com apenas alguns milhares de produtos, o número de combinações possíveis pode ser gigantesco e impossível de ser analisado e interpretado visualmente.

- **Perfil de produtos por loja** – como em todas as áreas em que se usa *analytics*, o processo começa com dados. O objetivo é integrar informações relevantes dentro da organização (estoque, vendas, perfis de loja, marketing e demografia). Em seguida, aplicando-se a analítica preditiva nesses dados, é possível avaliar perfis de loja e, em consequência, o abastecimento adequado de produtos para cada uma delas, com base em sua localização. Essa ferramenta permite, portanto, determinar a probabilidade de vendas de cada produto em determinada loja. O processo também possibilita medir o volume de vendas baseado em atributos associados a cada SKU. Podem ser levados em conta, fatores como lojas, *mixes* e, até mesmo, o clima. O resultado final é uma lista de SKUs a serem estocados em cada loja, ranqueados pela probabilidade de venda.
- **Fraude em ambientes financeiros** – os primeiros a utilizarem a análise de dados para prevenir fraude foram seguradoras, bancos e operadoras de telefonia. Uma das primeiras experiências com *analytics* bem-sucedida nessa prevenção no setor bancário é o sistema de identificação de fraudes *FICO Falcon*,



FIGURA 1 | CICLO COMPLETO DO ANALYTICS



baseado em redes neurais. Como o varejo também sofre com fraudes, sobretudo no ponto de venda, vários supermercados já utilizam redes de CFTV integradas aos PDVs. O aumento vertiginoso de compras via internet trouxe a reboque a fraude nas transações eletrônicas que, segundo estimativas, é doze vezes maior do que nas lojas físicas. Fraudes envolvendo telefonia celular, seguros, impostos, transações de cartões de crédito, dentre outros, representam problemas significativos para governos e empresas. No entanto, mesmo com todo o esforço envolvido, a detecção e a prevenção continuam sendo tarefas extremamente difíceis. Dizemos que a fraude é um crime adaptativo, isto é, procura estar sempre um passo à frente da detecção. Por isso, deve ser combatida com a utilização de métodos de análise de dados, mais ou menos avançados, utilizando KDD (*Knowledge Discovery in Databases*), *data mining* e *machine learning*.

#### AS QUATRO GRANDES CATEGORIAS DO ANALYTICS

O ciclo completo de *analytics* pode ser classificado em quatro grandes categorias: descritiva, diagnóstica, preditiva e prescritiva. As grandes questões que se impõem são: “o que está acontecendo?”, “por que está acontecendo?”, “quais as prováveis consequências?” e “o que posso fazer a respeito?” (Figura 1).

Como já citado, com a própria Wikipédia apontando a falta de clareza na diferenciação entre *analytics* e análise de dados, uma forma de

distingui-las talvez seja atribuir à análise de dados as tarefas relativas à compreensão do passado (analítica descritiva e diagnóstica) e, ao *analytics*, as tarefas de previsão do futuro (analítica preditiva e prescritiva).

**PRINCIPAIS ALGORITMOS DO ANALYTICS** Dentre os vários algoritmos utilizados em *analytics*, sete se destacam:

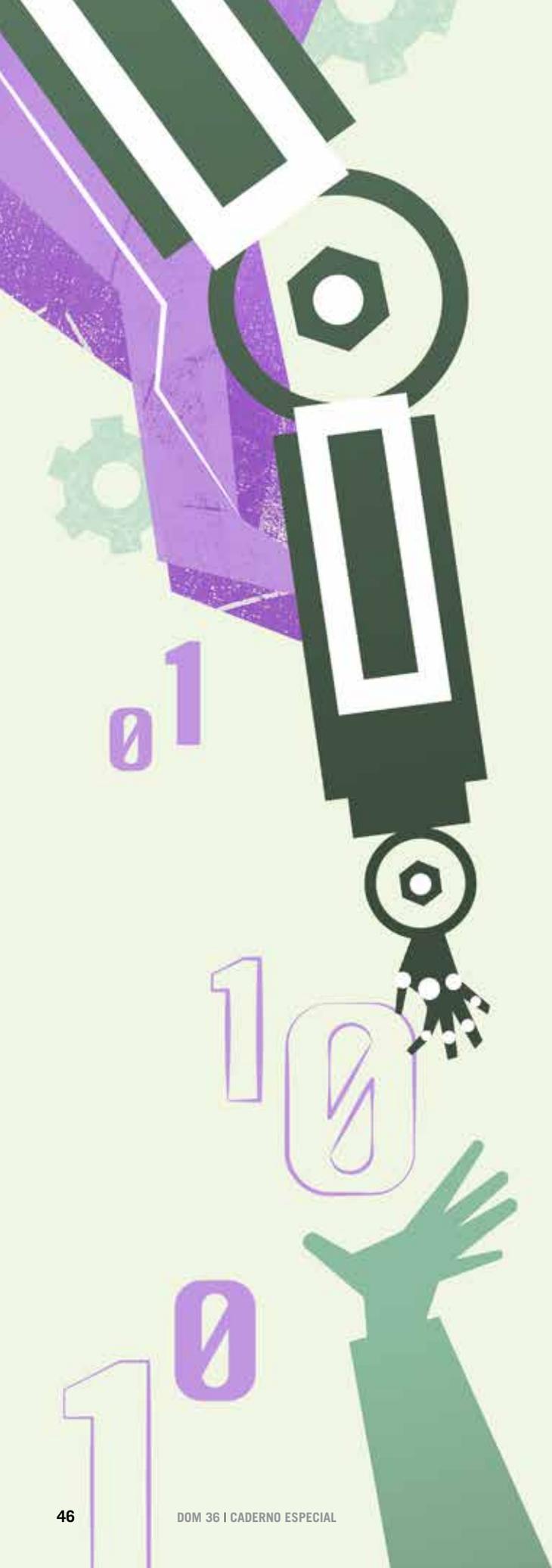
- **Teste de hipóteses** – apesar de não ser um algoritmo propriamente dito, é um conhecimento absolutamente necessário para qualquer cientista de dados. Consiste na utilização de testes estatísticos para verificar, com base nos dados existentes, se uma hipótese é verdadeira ou falsa. Baseados em testes hipotéticos, escolhe-se aceitar ou rejeitar uma hipótese. Quando um evento ocorre, pode ser uma tendência ou mero acaso. O teste de hipóteses dará a certeza de que foi uma ocorrência de importância, e não apenas um fato aleatório.
- **Regressão linear** – é uma técnica de modelagem estatística para avaliar o relacionamento entre uma variável explicativa e uma variável dependente, através do acondicionamento numa função linear dos pontos de dados observados. Um exemplo disso é a modelagem do índice de massa corporal, que utiliza o peso de uma amostragem de pessoas. A regressão linear é utilizada quando existe uma associação significativa entre variáveis. Com o mapeamento de uma quantidade expressiva de



## ATRIBUIR À ANÁLISE DE DADOS AS TAREFAS RELATIVAS À COMPREENSÃO DO PASSADO E, AO *ANALYTICS*, AS TAREFAS DE PREVISÃO DO FUTURO (ANALÍTICA PREDITIVA E PRESCRITIVA)

dados, será possível ter uma ideia se existe, ou não, associação clara entre as variáveis. Em caso negativo, a utilização da regressão linear provavelmente não será de muita utilidade.

- **Regressão logística** – como em outros tipos de regressão, essa técnica pode ser utilizada para avaliar o relacionamento entre variáveis de entrada e saída. No entanto, no caso de regressão logística, a variável de saída é sempre um resultado binário (0 ou 1, Sim ou Não, Verdadeiro ou Falso). Um exemplo típico é: haverá congestionamento na Marginal Tietê, nas proximidades da Ponte do Limão, amanhã, às 7h? O resultado é, necessariamente, sim ou não. A probabilidade de ocorrência de congestionamentos pode depender de atributos como clima, dia da semana, mês, hora do dia, número de veículos, etc. Usando a regressão logística podemos encontrar o modelo que melhor explique o relacionamento entre os atributos acima e as taxas de congestionamento de tráfego.
- **Técnicas de *clustering* (segmentação)** – é um tipo de algoritmo de aprendizado não supervisionado (*unsupervised learning*), no qual um conjunto de dados é agrupado em segmentos excludentes e diferenciados. Se, por exemplo, temos dados de clientes listados em 1.000 linhas, usando o *clustering* é possível agrupá-los em segmentos distintos, com base nas variáveis relativas a cada cliente, como idade, classe social, comportamento de compra, etc.



## O FATO É QUE A MAIORIA DAS ORGANIZAÇÕES JÁ UTILIZA *ANALYTICS* DE FORMA ROTINEIRA E CONSISTENTE, COM O OBJETIVO DE MELHORAR O DESEMPENHO

- **ANOVA** – *Analysis of Variance* (Análise de Variância) – teste utilizado para determinar se as médias de mais de dois grupos de dados são significativamente distintas entre si. Por exemplo, quando iniciamos uma campanha do tipo “compre um e ganhe outro”, destinada a cinco grupos de 100 clientes cada, de diferentes atributos demográficos. Supondo que conseguíssemos realmente verificar que cada grupo respondeu de forma diferente à campanha, poderíamos, então, criar uma promoção voltada para cada grupo específico, reduzindo o custo total da campanha e aumentando a taxa de sucesso. A análise permite comparar tanto a variação entre os grupos distintos quanto a de todo o grupo. A essência dessa técnica consiste em avaliar se todos os grupos fazem parte de uma população maior ou se possuem características totalmente distintas.
- **Redes neurais** – também conhecidas como redes neurais artificiais, se inspiram no sistema nervoso humano. Mais precisamente, em como as informações complexas são absorvidas e processadas pelo sistema. Assim como os seres humanos, as redes neurais também aprendem através de experiências e exemplos. Elas são sempre configuradas para uma aplicação específica e utilizadas para identificar padrões dentro de conjuntos complexos de dados e, assim, fornecer previsões e classificar dados. Normalmente, são organizadas em camadas, compostas por determinado número de “nós” interconectados. Os padrões são apresentados à rede pela camada de dados de entrada, que se comunica a uma ou mais “camadas ocultas”, onde o processamento efetivamente ocorre. As “camadas ocultas” se comunicam, então, com a camada de dados de saída (**Figura 2**).
- **Árvores de decisão** – como o próprio nome sugere,

é a representação, em forma de árvore, de como chegar a uma decisão específica por meio do mapeamento de todas as opções existentes e suas probabilidades de ocorrência. São extremamente fáceis de ser compreendidas e interpretadas. Podemos verificar, em cada “nó” da árvore, qual seria a consequência de selecionar aquela opção.

---

MIGUEL VILLELA é especialista em *analytics*.

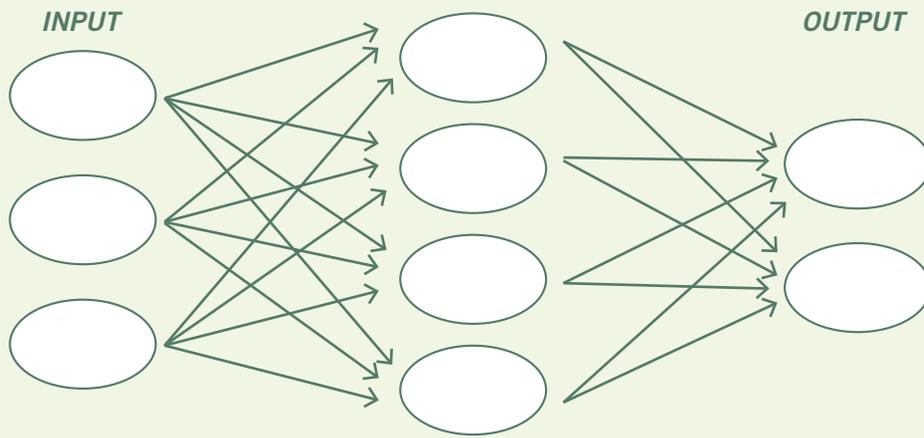
---

AGOSTINHO VILLELA é Líder do Processo de Inovação para a América Latina e membro da Academia de Tecnologia da IBM.

---

HUGO TADEU é professor do Núcleo de Inovação e Empreendedorismo da Fundação Dom Cabral.

FIGURA 2 | EXEMPLO DE UMA REDE NEURAL



### CONCLUSÃO

Num mundo em que o volume de dados armazenados já é enorme e só faz crescer a cada dia, não importa se ainda não temos uma definição clara de *analytics* ou das diferenças que existem entre ele e *data science*, *data mining*, *machine learning*, etc. O fato é que a maioria das organizações já utiliza *analytics* de forma rotineira e consistente, com o objetivo de melhorar o seu desempenho. Essa nova inteligência possibilita uma melhor compreensão dos dados e amplia a capacidade de previsão da empresa em seu mercado de atuação, além ser imprescindível na tomada de decisões.

### PARA SE APROFUNDAR NO TEMA

GUPTA, Bhasker. 10 machine learning algorithms every data scientist should know. **Analytics India Magazine**. 2017. Disponível em: <<https://analyticsindiamag.com/10-machine-learning-algorithms-every-data-scientist-know/>>. Acesso em: 28 jan. 2019.

INTERNATIONAL BUSINESS MACHINES. **Store assortment analytics**. 2010. Disponível em: <[http://public.dhe.ibm.com/software/data/sw-library/cognos/pdfs/solutionbriefs/sb\\_store\\_assortment\\_analytics.pdf](http://public.dhe.ibm.com/software/data/sw-library/cognos/pdfs/solutionbriefs/sb_store_assortment_analytics.pdf)>. Acesso em: 28 jan. 2019.

RETAIL WIRE. **Building a data-driven pricing strategy**. Disponível em: <[http://www.wns.com/Portals/0/Documents/Whitepapers/PDFFiles/651/35/WNS\\_Whitepaper\\_Pricing-Analytics.pdf](http://www.wns.com/Portals/0/Documents/Whitepapers/PDFFiles/651/35/WNS_Whitepaper_Pricing-Analytics.pdf)>. Acesso em: 28 jan. 2019.

SYSTEM DESIGN MANAGEMENT. **Analytics for decision making**. Disponível em: <<https://sdm.mit.edu/systemsthinkingconference/2013/presentations/catanzaro.pdf>>. Acesso em: 28 jan. 2019.