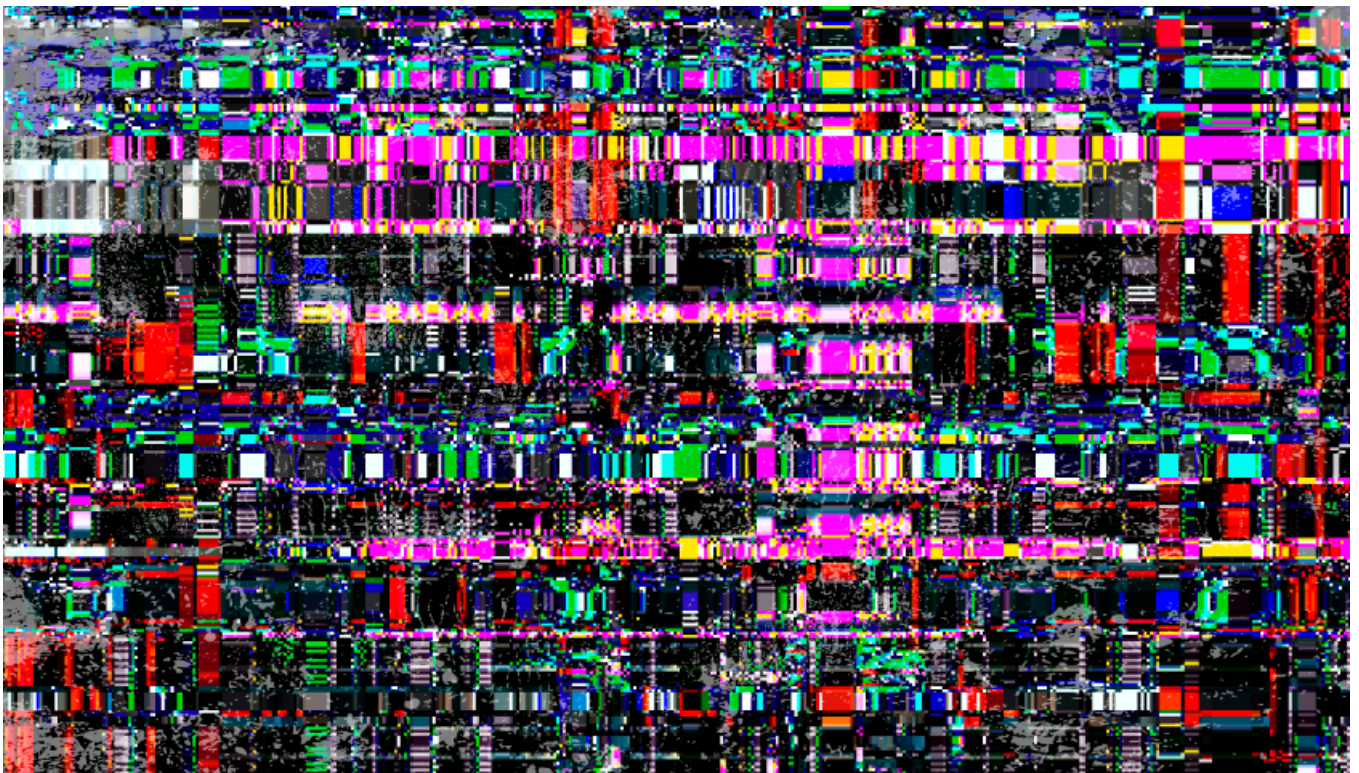


# Por que tantos projetos de ciência de dados fracassam

Organizações podem se beneficiar mais das análises sofisticadas identificando e superando cinco obstáculos frequentes



Organizações podem se beneficiar mais das análises sofisticadas identificando e superando cinco obstáculos frequentes

Cada vez mais empresas estão adotando a ciência de dados como atividade ou competência, mas muitas não têm conseguido obter resultados consistentes de seus investimentos em big data, inteligência artificial ou machine learning. Pode-se até dizer que é crescente a distância entre as que usufruem dos benefícios gerados daquelas que ainda lutam para alcançar resultados positivos.

Este artigo se baseia em uma pesquisa sobre iniciativas de ciência de

dados de três bancos privados indianos, cujos ativos somam mais de US\$ 200 milhões. O estudo incluiu observações in loco, entrevistas semiestruturadas com 57 executivos, gestores e cientistas de dados, além da análise de registros e documentos. Nossa pesquisa identificou os cinco erros de abordagem mais comuns quando se trata de ciência de dados, como se verá a seguir, com sugestões de como evitá-los, segundo os exemplos coletados.

## **Caso 1: Um martelo à procura de um prego**

Esta é uma situação bastante comum, em que se formula uma solução e, a partir dela, busca-se um problema em que possa ser aplicada.

Hiren, cientista de dados recém-contratado por um dos bancos pesquisados, é um tipo de analista de dados cobiçado por muitas empresas. Conhece profundamente o algoritmo KNN (K-Nearest Neighbors) usado em problemas de classificação de dados em machine learning. Ele estava ansioso por aplicá-lo no “mundo real” para a identificação e a classificação de clusters de dados e o fez pensando em identificar segmentos da economia especialmente lucrativos a partir da avaliação do portfólio de correntistas do banco.

Hiren decidiu usar esse algoritmo para analisar os segmentos da economia dos quais os clientes do banco faziam parte e, a partir dos resultados, recomendou que fossem privilegiados dois dos 33 segmentos existentes no grupo, porque teriam resultados mais promissores. Tal conclusão causou perplexidade nos gerentes de contas e analistas que atendiam a esses segmentos, uma vez que eles próprios, pela observação diária, já haviam chegado a conclusão semelhante, sem precisar de um recurso tão sofisticado.

Nesse caso e em outros três bancos analisados, as falhas em se identificar corretamente oportunidades de negócios rentáveis decorreu da paixão por soluções baseadas em ciência de dados. Essa falha pode

se manifestar de várias maneiras. No caso de Hiren, o problema não exigia uma solução tão elaborada. Em outras situações, vimos o uso bem-sucedido de uma solução em determinado segmento ser usado como justificativa para seu uso em outro, no qual não era tão apropriado ou eficaz. Em suma, as falhas não decorrem da incorreta execução das técnicas analíticas e, sim, de sua má aplicação.

Ao se familiarizar mais com o negócio, Hiren propôs usar novamente o algoritmo KNN, mas, em vez de aplicá-lo comparando as indústrias, usou-o para clusterizar os clientes. Foi quando ele conseguiu identificar novos insights que permitiram à equipe de gestão de contas identificar grupos de clientes ainda inexplorados e com elevado potencial para operações lucrativas.

Não é preciso muito para ver que soluções analíticas funcionam melhor quando levam em consideração o negócio, e não contrário. Porém, muitos ainda se deslumbram com a tecnologia, particularmente se ela vier com rótulos que impressionam, como inteligência artificial ou machine learning. Ao mesmo tempo, cientistas de dados muitas vezes não conseguem deixar clara essa perspectiva, de que as ferramentas não são solução em si mesmas.

Para enfrentar o problema, executivos dos bancos pesquisados têm investido em treinamento. Em um deles, cientistas de dados fizeram cursos sobre os produtos bancários junto com os trainees da área gerencial. Da mesma forma, todo o pessoal de atendimento recebeu um treinamento em ciência de dados formulado especialmente para eles, abordando os conceitos básicos, com ênfase em como solicitar soluções mais adequadas às demandas do negócio. Essa ação aumentou a integração das equipes e incrementou o conhecimento dos respectivos campos de conhecimento.

Adicionalmente, vimos também a formulação de processos para evitar a tentação de adotar soluções tecnológicas prematuramente. Temos o

exemplo de uma companhia aeroespacial norte-americana que usa uma metodologia chamada Seven Ways, que exige que se comparem ao menos sete possíveis soluções e se justifique a escolha por uma delas.

## **Caso 2: Detalhes desconsiderados**

Pranav, cientista de dados especialista em modelagem estatística, desenvolveu um algoritmo para facilitar a concessão de empréstimos para pequenas e médias empresas, com o objetivo de facilitar o trabalho dos analistas de crédito. Usando os registros de créditos aprovados dos dez anos anteriores, ele comparou a saúde financeira dos tomadores de empréstimos no momento da concessão com a atual. Surgia uma ferramenta de software baseada em um modelo de alta precisão, que a área de crédito prontamente implantou. Porém, passados seis meses, ficou claro que a taxa de inadimplência havia crescido, comparativamente aos empréstimos concedidos antes da ferramenta. Espantados, um analista de crédito experiente foi designado para ajudar a descobrir onde estava o erro.

Não demorou para encontrarem o motivo: toda a análise havia sido feita levando em consideração somente os registros de empréstimos aprovados após a análise dos gerentes de conta, sem que se usasse qualquer informação dos casos rejeitados. Com isso, o modelo supunha que todos os clientes tinham a mesma qualidade, o que não era verdade. Ao incluir as estatísticas negativas, o software passou a funcionar como se esperava.

O maior problema para as empresas que querem capturar valor com a ciência de dados é como discernir as fontes de vieses com antecedência, e garantir que isso não seja replicado nos modelos. Isso é desafiador porque pessoas leigas – e às vezes até os especialistas em analytics – não conseguem dizer com facilidade como os dados e o analytics geram os resultados. E os especialistas que entendem nem

sempre reconhecem os vieses ocultos nos dados brutos que utilizam.

Os bancos que participaram da pesquisa procuram evitar esse viés aproximando os cientistas de dados dos analistas de negócios, expondo-os a situações variadas em que possam identificar os dados a serem considerados. Também pudemos observar um grupo que se utiliza de um processo de depuração dos dados, de forma a minimizar análises enviesadas.

### **Caso 3: Solução correta, momento errado**

Kartik, profissional especialista em machine learning, levou um mês para desenvolver um sistema que analisasse a perda de clientes de contas poupança e mais três afinando esses dados. Ao apresentar a solução aos gestores do produto, já não havia mais recursos para a implantação. No afã de evitar que isso se repetisse no ano seguinte, Kartik reapresentou seu modelo à mesma equipe a tempo de incluir no orçamento do ano seguinte, mas tudo se repetiu: a prioridade deixou de ser a manutenção de clientes e tornou-se a conquista de novos, não havendo razão para usar o software. No terceiro ano, o interesse no modelo e o orçamento se encontraram, mas aí o próprio cientista de dados considerou que o projeto estava ultrapassado e precisaria ser refeito.

A falta de sincronia entre a área de desenvolvimento tecnológico e as prioridades do negócio impede, em muitos casos, que as empresas tirem proveito das tecnologias de ponta. A direção das empresas precisa garantir que haja uma integração efetiva entre a estratégia da organização e a ciência de dados, em termos físicos, estruturais e de processos. Foi o que fez um dos bancos com os quais trabalhamos, que inseriu os cientistas de dados nas equipes de negócios a cada projeto. Isso permitiu a convivência entre as equipes, de forma que as prioridades e as necessidades se tornassem conhecidas por todos.

Resumidamente, especialistas em dados devem concentrar seus

esforços nas questões prioritárias de acordo com a estratégia definida pela organização. No entanto, duas ressalvas devem ser feitas:

Se o profissional de ciência de dados tiver uma ideia, um insight que faça sentido para o negócio, este deve, sim, ser apresentado e analisado – isso não é privilégio das áreas de negócios. Cientistas de dados alocados em um time da linha de frente estão a serviço da organização como um todo. As soluções devem ser parte de um projeto, não caprichos setoriais.

## **Caso 4: Solução certa, usuário errado**

Sophia, analista de negócios de um dos bancos que participaram do estudo, trabalhou com sua equipe no desenvolvimento de um motor para otimizar recomendações de produtos e serviços para clientes. Uma vez concluído, o modelo foi incluído em todas as plataformas digitais de relacionamento com os clientes e esperava-se uma forte adesão. No entanto, isso não acontecia. Fez-se, então, uma pesquisa por telefone com uma amostra de clientes e o mistério foi rapidamente solucionado: muitos clientes não confiavam em recomendações feitas por canais digitais.

Para verificar com mais precisão o que estava havendo, Sophia visitou várias agências e viu que os clientes aceitavam muito bem as recomendações de investimentos e serviços feitas pelos gerentes de relacionamento. Assim, mesmo que o sistema encontrasse as melhores sugestões, elas ganhavam força com o aval desses gerentes. A solução foi simples: o motor de busca passou a ser usado pelos gerentes de relacionamento que atendiam pessoas físicas e jurídicas com sucesso.

Por melhor que seja o resultado de uma ferramenta analítica, a forma como seus resultados são passados adiante deve ser levada em consideração. Isso inclui avaliar que usuários mais se beneficiarão, até por meio de testes. E, embora até a conclusão de nossa pesquisa não

tenhamos encontrado essa prática (testes feitos por pessoas), pudemos ver exemplos como o caso relatado, em que o modelo acabou positivamente encoberto pela ação dos gerentes de conta. O uso de uma interface (tecnológica ou não) que atenda às necessidades do usuário final contribui para que os resultados gerados pela ciência de dados sejam úteis e bem aceitos.

## **Caso 5: O final turbulento**

A tentativa do banco de recuperar clientes estava parada há meses, e a reunião entre cientistas de dados, Dhara e Viral, e gerentes de produto, Anish e Jalpa, que deveria colocar a iniciativa nos trilhos, também não ia bem. Os primeiros se concentravam na busca de motivos para os clientes voltarem, enquanto os demais focavam em detalhes da campanha a ser feita e pressionavam o pessoal de TI a implantar a solução imediatamente. Sem chegarem a um acordo, Viral comentou: “se cientistas de dados e analistas fazem tudo, para que gerentes de produto? Nosso papel é desenvolver uma solução analítica, eles que executem”.

Felizmente, os cientistas de dados reviram sua posição e entenderam a necessidade de implantação, fruto de interações passadas com sistemas que não funcionavam, apenas produziam listas de ex-clientes com dados ultrapassados, para que os gerentes de relacionamento trabalhassem a partir delas. A solução foi criar uma etapa intermediária, em que telemarketing ativo, e-mails e profissionais das agências teriam à disposição um software para fazer a atualização da base de dados, para então identificar listas de clientes a serem recuperados que pudessem realmente auxiliar no trabalho dos gerentes de produto. Só então o projeto realmente vingou.

O que se depreende é que o episódio exigiu de Dhara e Viral uma inesperada dose de empatia e iniciativa, assumindo o papel de líderes do

projeto e criando novas etapas, ainda que inicialmente isso não estivesse nos planos. Afinal, cientistas de dados ainda são relativamente raros e caros, e não deveriam assumir funções fora de sua área; porém, sem isso, o projeto teria fracassado.

Sim, cientistas de dados são raros e caros, mas isso não significa que se deva “economizá-los”. As empresas podem e devem envolvê-los nos processos de implantação de soluções. Nosso estudo verificou isso ao perceber que eles se sentiam motivados quando asseguravam o sucesso das soluções que criavam e participavam de sua efetivação. De fato, alguns executivos reconheciam que esses profissionais ultrapassavam muito as fronteiras de seus cargos, mas que isso pode ser administrado caso a caso, mantendo seu envolvimento com o negócio.

TODOS OS PROBLEMAS QUE SURGIRAM AO LONGO DE NOSSA PESQUISA SÃO, INVARIAVELMENTE, FRUTO INTERAÇÕES NEGATIVAS ENTRE CIÊNCIA DE DADOS E GESTÃO. A liderança deve incentivar uma ampliação do papel da ciência de dados em suas empresas, de modo a integrar o diagnóstico de problemas, a gestão de processos e a implantação de soluções. Isso pode ocorrer por meio de treinamentos, da criação de interfaces que facilitam a interação do usuário ou mesmo de incentivos formais. O resultado é a queda no número de soluções mal sucedidas, a implantação mais rápida de projetos e maior retorno nos negócios.