

# O que é Ciência de Dados?

A ciência de dados combina vários campos, incluindo estatísticas, métodos científicos, inteligência artificial (IA) e análise de dados para extrair valor dos dados. Aqueles que praticam a ciência de dados são chamados de cientistas de dados e combinam uma variedade de habilidades para analisar dados coletados da web, smartphones, clientes, sensores e outras fontes para obter insights acionáveis.

A ciência de dados abrange a preparação de dados para análise, incluindo limpeza, agregação e manipulação de dados para realizar análises de dados avançadas. Os aplicativos analíticos e os cientistas de dados podem então revisar os resultados para descobrir padrões e permitir que os líderes de negócios obtenham insights informados.

## **Ciência de dados: Um recurso inexplorado para machine learning**

Ciência de dados é um dos campos mais interessantes que existem hoje. Por que é tão importante?

Porque as empresas estão sentadas em um tesouro de dados. Como a tecnologia moderna permitiu a criação e armazenamento de quantidades crescentes de informações, os volumes de dados explodiram. Estima-se que 90% dos dados do mundo tenham sido criados nos últimos dois anos. Por exemplo, os usuários do Facebook carregam 10 milhões de fotos a cada hora.

Mas esses dados muitas vezes ainda estão apenas em bancos de dados e data lakes, em sua maioria intocados.

A riqueza de dados coletados e armazenados por essas tecnologias pode trazer benefícios transformadores para organizações e sociedades em todo o mundo, mas somente se pudermos interpretá-los. É aí que entra a ciência de dados.

A ciência de dados revela tendências e produz as informações que as empresas podem usar para tomar melhores decisões e criar produtos e serviços mais inovadores. Talvez o mais importante seja que ele permite que os modelos de machine learning (ML) aprendam com as grandes quantidades de dados que estão sendo fornecidos a eles, em vez de depender

Os dados são a base da inovação, mas seu valor vem dos dados de informações que os cientistas podem extrair e depois usar.

## Qual é a diferença entre ciência de dados, inteligência artificial e machine learning?

Para entender melhor a ciência de dados - e como você pode aproveitá-la - é igualmente importante conhecer outros termos relacionados à área, como inteligência artificial (IA) e machine learning. Frequentemente, você descobrirá que esses termos são usados de forma intercambiável, mas existem nuances.

Aqui está uma análise simples:

IA significa fazer com que um computador imite o comportamento humano de alguma forma.

Ciência de dados é um [subconjunto da inteligência artificial \(IA\)](#) e se refere mais às áreas sobrepostas de estatísticas, métodos científicos e análise de dados - todas as quais são usadas para extrair significado e percepções dos dados.

[Machine learning](#) é outro subconjunto da inteligência artificial (IA) e consiste nas técnicas que permitem aos computadores descobrir coisas a partir dos dados e fornecer aplicativos de inteligência artificial (IA).

E para garantir, vamos adicionar outra definição.

Aprendizagem profunda (deep learning), que é um subconjunto de machine learning que permite aos computadores resolver problemas mais complexos.

## Como a ciência de dados está transformando os negócios

As organizações estão usando equipes de ciência de dados para transformar os dados em uma vantagem competitiva, refinando produtos e serviços. Os casos de uso de ciência de dados e machine learning incluem:

Determinar a rotatividade de clientes analisando os dados coletados nos call centers, para que o marketing possa tomar medidas para retê-los

Melhore a eficiência analisando padrões de tráfego, condições meteorológicas e outros fatores para que as empresas de logística possam melhorar as velocidades de entrega e reduzir custos

Melhore o diagnóstico de pacientes analisando dados de exames médicos e sintomas relatados para que os médicos possam diagnosticar doenças mais cedo e tratá-las com mais eficácia

Otimize a cadeia de abastecimento, prevendo quando o equipamento irá quebrar

ações anômalas

Melhore as vendas criando recomendações para clientes com base em compras anteriores

Muitas empresas priorizaram a ciência de dados e estão investindo fortemente nela. [Na pesquisa recente do Gartner](#) com mais de 3.000 Diretores de Informações (CIOs), os entrevistados classificaram o business intelligence e a análise avançada como o principal diferencial de tecnologia para as suas empresas. Os Diretores de Informações (CIOs) entrevistados consideram essas tecnologias as mais estratégicas para suas empresas e estão investindo de acordo.

## Como a ciência de dados é conduzida

O processo de análise e ação sobre os dados é iterativo em vez de linear, mas é assim que o ciclo de vida da ciência de dados normalmente flui para um projeto de modelagem de dados:

**Planejamento:** Definir um projeto e seus possíveis resultados.

**Construção de um modelo de dados:** Os cientistas de dados costumam usar uma variedade de bibliotecas de código aberto ou ferramentas no banco de dados para construir modelos de machine learning. Frequentemente, os usuários desejam APIs para ajudar na ingestão de dados, criação de perfil de dados e visualização ou engenharia de recursos. Eles precisarão das ferramentas certas, assim como de acesso aos dados certos e a outros recursos, como poder de computação.

**Avaliação de um modelo:** Os cientistas de dados devem atingir um alto percentual de precisão para seus modelos antes de se sentirem confiantes em implementá-los. A avaliação do modelo normalmente irá gerar um conjunto abrangente de métricas de avaliação e visualizações para medir o desempenho do modelo em relação aos novos dados e também classificá-los ao longo do tempo para permitir o comportamento ideal na produção. A avaliação do modelo vai além do desempenho bruto para levar em consideração o comportamento esperado da linha de base.

**Explicação de modelos:** Ser capaz de explicar a mecânica interna dos resultados dos modelos de machine learning em termos humanos nem sempre foi possível - mas está se tornando cada vez mais importante. Os cientistas de dados desejam explicações automatizadas do peso relativo e da importância dos fatores que geram uma previsão e detalhes explicativos específicos do modelo nas previsões do modelo.

**Implementação de um modelo:** Usar um modelo de machine learning treinado e colocá-lo nos sistemas certos costuma ser um processo difícil e trabalhoso. Isso pode ser facilitado operacionalizando modelos como APIs escaláveis e seguras ou usando modelos de machine learning no banco de dados.

**Monitoramento de modelos:** Infelizmente, implementar um modelo não é o fim disso. Os modelos devem sempre ser monitorados após a implementação para garantir que estejam funcionando corretamente. Os dados em que o modelo foi treinado podem não ser mais

fraudes, os criminosos estão sempre inventando novas maneiras de hackear contas.

## Ferramentas para ciência de dados

Criar, avaliar, implementar e monitorar modelos de machine learning pode ser um processo complexo. É por isso que houve um aumento no número de ferramentas de ciência de dados. Os cientistas de dados usam muitos tipos de ferramentas, mas uma das mais comuns são os notebooks de código aberto, que são aplicativos da web para gravar e executar códigos, visualizar dados e ver os resultados, tudo no mesmo ambiente.

Alguns dos notebooks mais populares são Jupyter, RStudio e Zeppelin. Os notebooks são muito úteis para conduzir análises, mas têm suas limitações quando os cientistas de dados precisam trabalhar em equipe. Plataformas de ciência de dados foram criadas para resolver esse problema.

Para determinar qual ferramenta de ciência de dados é certa para você, é importante fazer as seguintes perguntas: Que tipo de linguagem seus cientistas de dados usam? Que tipo de método de trabalho eles preferem? Quais tipos de fontes de dados eles estão usando?

Por exemplo, alguns usuários preferem ter um serviço independente de fonte de dados que usa bibliotecas de código aberto. Outros preferem a velocidade dos algoritmos de machine learning no banco de dados.

Experimente um workshop de machine learning.

## Quem supervisiona o processo de ciência de dados?

Na maioria das organizações, os projetos de ciência de dados são normalmente supervisionados por três tipos de gerentes:

**Gerentes de negócios:** Esses gerentes trabalham com a equipe de ciência de dados para definir o problema e desenvolver uma estratégia para análise. Eles podem ser o chefe de uma linha de negócios, como marketing, finanças ou vendas, e ter uma equipe de ciência de dados se reportando a eles. Eles trabalham em estreita colaboração com os gerentes de ciência de dados e TI para garantir que os projetos sejam entregues.

**Gerentes de TI:** Os gerentes seniores de TI são responsáveis pela arquitetura e pelo planejamento da infraestrutura que oferecerão suporte para as operações de ciência de dados. Eles monitoram continuamente as operações e o uso de recursos para garantir que as equipes de ciência de dados operem de forma eficiente e segura. Esses gerentes também podem ser responsáveis pela criação e atualização de ambientes de TI para equipes de ciência de dados.

**Gerentes de ciência de dados:** Esses gerentes supervisionam a equipe de ciência de dados e seu trabalho diário. Eles são criadores de equipes que podem equilibrar o desenvolvimento da equipe com o planejamento e o monitoramento do projeto.

## O que é um cientista de dados?

Como especialidade, a ciência de dados é jovem. Ela cresceu a partir dos campos de análise estatística e mineração de dados (data mining). O *Data Science Journal* estreou em 2002, publicado pelo Conselho Internacional para Ciências: Comitê de Dados para Ciências e Tecnologia. Em 2008, o título de cientista de dados surgiu e a área decolou rapidamente. Houve uma escassez de cientistas de dados desde então, embora mais e mais faculdades e universidades tenham começado a oferecer graduação em ciência de dados.

As funções de um cientista de dados podem incluir o desenvolvimento de estratégias para analisar dados, preparar dados para análise, explorar, analisar e visualizar dados, construir modelos com dados usando linguagens de programação, como Python e R, e implementar modelos em aplicativos.

O cientista de dados não trabalha sozinho. Na verdade, a ciência de dados mais eficaz é feita em equipes. Além de um cientista de dados, essa equipe pode incluir um analista comercial que define o problema, um engenheiro de dados que prepara os dados e como eles são acessados, um arquiteto de TI que supervisiona os processos e a infraestrutura subjacentes e um desenvolvedor de aplicativos que implementa os modelos ou os resultados da análise em aplicativos e produtos.

## Desafios de implementação de projetos de ciência de dados

Apesar da promessa da ciência de dados e dos enormes investimentos em equipes de ciência de dados, muitas empresas não estão percebendo o valor total de seus dados. Em sua corrida para contratar talentos e criar programas de ciência de dados, algumas empresas experimentaram fluxos de trabalho de equipe ineficientes, com pessoas diferentes usando diferentes ferramentas e processos que não funcionam bem juntos. Sem um gerenciamento mais disciplinado e centralizado, os executivos podem não ver o retorno total de seus investimentos.

Esse ambiente caótico apresenta muitos desafios.

Cientistas de dados não podem trabalhar com eficiência. Como o acesso aos dados deve ser concedido por um administrador de TI, os cientistas de dados costumam esperar muito tempo pelos dados e pelos recursos necessários para analisá-los. Assim que obtém acesso, a equipe de ciência de dados pode analisar os dados usando ferramentas diferentes e possivelmente incompatíveis. Por exemplo, um cientista pode desenvolver um modelo usando a linguagem R, mas o aplicativo em que será usado é escrito em uma linguagem diferente. É por isso que pode levar semanas, ou mesmo meses, para implementar os modelos em aplicativos úteis.

Os desenvolvedores de aplicativos não podem acessar o machine learning utilizável. Às vezes, os modelos de machine learning que os desenvolvedores recebem precisam ser

pontos de acesso podem ser inflexíveis, os modelos não podem ser implementados em todos os cenários e a escalabilidade é deixada para o desenvolvedor do aplicativo.

Os administradores de TI gastam muito tempo em suporte. Por causa da proliferação de ferramentas de código aberto, a TI pode ter uma lista cada vez maior de ferramentas para oferecer suporte. Um cientista de dados em marketing, por exemplo, pode estar usando ferramentas diferentes de um cientista de dados em finanças. As equipes também podem ter fluxos de trabalho diferentes, o que significa que a equipe de TI deve reconstruir e atualizar continuamente os ambientes.

Os gerentes de negócios estão muito distantes da ciência de dados. Os fluxos de trabalho de ciência de dados nem sempre são integrados aos sistemas e processos de tomada de decisões de negócios, dificultando a colaboração dos gerentes de negócios de maneira conhecida com os cientistas de dados. Sem uma melhor integração, os gerentes de negócios acham difícil entender por que leva tanto tempo para ir do protótipo à produção, e é menos provável que eles apoiem o investimento em projetos que acreditam ser lentos demais.

Saiba mais sobre o ciclo de vida da ciência de dados (PDF)

## A plataforma de ciência de dados oferece novos recursos

Muitas empresas perceberam que, sem uma plataforma integrada, o trabalho de ciência de dados era ineficiente, inseguro e difícil de dimensionar. Essa percepção levou ao desenvolvimento de [plataformas de ciência de dados](#). Essas plataformas são hubs de software em torno dos quais todo o trabalho de ciência de dados ocorre. Uma boa plataforma alivia muitos dos desafios da implementação de ciência de dados e ajuda as empresas a transformar seus dados em informações de maneira mais rápida e eficiente.

Com uma plataforma de machine learning centralizada, os cientistas de dados podem trabalhar em um ambiente colaborativo usando suas ferramentas de código aberto favoritas, com todo o seu trabalho sincronizado por um sistema de controle de versão.

## Os benefícios de uma plataforma de ciência de dados

Uma plataforma de ciência de dados reduz a redundância e impulsiona a inovação, permitindo que as equipes compartilhem códigos, resultados e relatórios. Ela remove gargalos no fluxo de trabalho, simplificando o gerenciamento e incorporando as boas práticas.

Em geral, as melhores plataformas de ciência de dados visam:

- Tornar os cientistas de dados mais produtivos, ajudando-os a acelerar e entregar modelos mais rapidamente e com menos erros

- Facilitar o trabalho dos cientistas de dados com grandes volumes e variedades de dados

- Fornecer inteligência artificial confiável de nível empresarial que é livre de preconceitos,

As plataformas de ciência de dados são construídas para a colaboração de uma variedade de usuários, incluindo cientistas de dados especialistas, [cientistas de dados do cidadão](#), [engenheiros de dados](#) e engenheiros ou especialistas em machine learning. Por exemplo, uma plataforma de ciência de dados pode permitir que cientistas de dados implementem modelos como APIs, facilitando sua integração em diferentes aplicativos. Os cientistas de dados podem acessar ferramentas, dados e infraestrutura sem ter que esperar pela equipe de TI.

A demanda por plataformas de ciência de dados explodiu no mercado. De fato, [o mercado de plataformas deverá crescer](#) a uma taxa anual composta de mais de 39% nos próximos anos e está projetado para atingir US\$ 385 bilhões até 2025.

## O que um cientista de dados precisa em uma plataforma

Se você estiver pronto para explorar os recursos das plataformas de ciência de dados, há alguns recursos importantes a serem considerados:

Escolha uma interface do usuário baseada em projeto que incentive a colaboração. A plataforma deve capacitar as pessoas a trabalharem juntas em um modelo, desde a concepção até o desenvolvimento final. Ela deve fornecer a cada membro da equipe acesso de autoatendimento aos dados e recursos.

Priorize a integração e a flexibilidade. Certifique-se de que a plataforma inclua suporte para as ferramentas de código aberto mais recentes, provedores de controle de versão comuns, como GitHub, GitLab e Bitbucket e forte integração com outros recursos.

Inclua recursos de nível empresarial. Assegure-se de que a plataforma possa escalar com seus negócios à medida que sua equipe cresce. A plataforma deve estar altamente disponível, ter controles de acesso robustos e suportar um grande número de usuários simultâneos.

Torne a ciência de dados mais autônoma. Procure uma plataforma que tire o peso da equipe de TI e da engenharia e facilite para os cientistas de dados criarem ambientes instantaneamente, acompanharem todo o trabalho e implementarem modelos facilmente na produção.

Garanta uma implementação de modelo mais fácil. A implementação e a operacionalização do modelo são uma das etapas mais importantes do ciclo de vida de machine learning, mas costuma ser desconsiderada. Certifique-se de que o serviço escolhido facilite a operacionalização de modelos, seja fornecendo APIs ou garantindo que os usuários criem modelos de uma forma que permita uma integração fácil.

## Quando uma plataforma de ciência de dados é a medida certa

Sua organização pode estar pronta para uma plataforma de ciência de dados, se você percebeu

Produtividade e colaboração estão mostrando sinais de tensão

Modelos de machine learning não podem ser auditados ou reproduzidos

Modelos nunca chegam à produção

Uma plataforma de ciência de dados pode agregar valor real ao seu negócio. A [plataforma de ciência de dados](#) da Oracle inclui uma ampla gama de serviços que fornecem uma experiência abrangente de ponta a ponta, projetada para acelerar a implementação do modelo e melhorar os resultados de ciência de dados.



## Oracle Cloud - Modo Gratuito (Free Tier)

Crie, teste e implemente aplicativos aplicando processamento de linguagem natural - gratuitamente.

[Inscreva-se agora](#)

## Tópicos de ciência de dados

[Ciência de dados: Um recurso inexplorado para machine learning](#)

[Qual é a diferença entre ciência de dados, inteligência artificial e machine learning?](#)

[Como a ciência de dados está transformando os negócios](#)

[Como a ciência de dados é conduzida](#)

[Ferramentas para ciência de dados](#)

[Quem supervisiona o processo de ciência de dados?](#)

[O que é um cientista de dados?](#)

[Desafios de implementação de projetos de ciência de dados](#)

[A plataforma de ciência de dados oferece novos recursos](#)

[Os benefícios de uma plataforma de ciência de dados](#)

[O que um cientista de dados precisa em uma plataforma](#)

[Quando uma plataforma de ciência de dados é a medida certa](#)

Recursos para

Carreiras  
Desenvolvedores

Por que Oracle

Relatórios de  
Analistas  
Gartner MQ para

Aprenda

O que é  
computação em  
nuvem?

Novidades

Experimente o  
Oracle Cloud -  
Modo Gratuito

Parceiros  
Startups  
Estudantes e Educadores

Economia da Nuvem  
Responsabilidade Corporativa  
Diversidade e Inclusão  
Práticas de Segurança

O que é Docker?  
O que é Kubernetes?  
O que é Python?  
O que é SaaS?

Oracle Arm Processors  
Oracle e Premier League  
Oracle e Red Bull Racing Honda  
Plataforma de Experiência do Funcionário  
Oracle Support Rewards

## Entre em Contato

Vendas: 0800-891-4433

Como podemos ajudar?

Inscreva-se para receber e-mails

Eventos

Notícias

Blogs

País/Região

© 2021 Oracle

Mapa do Site

Termos de Uso e Privacidade

Preferências de Cookies

Opções de Anúncios

Carreiras