

Brought to you by:



# Machine Learning

for  
**dummies**<sup>®</sup>  
A Wiley Brand

Understand machine learning fundamentals

Make sense of machine learning algorithms

Build your data science team



**Judith Hurwitz**  
**Daniel Kirsch**

**IBM Limited Edition**



# Aprendizado de máquina

IBM Edição Limitada

por Judith Hurwitz e  
Daniel Kirsch

**for  
dummies<sup>®</sup>**  
A Wiley Brand

## Aprendizado de Máquina For Dummies®, IBM Edição Limitada

Publicado por  
**John Wiley & Sons, Inc.**  
111 Rio St.  
Hoboken, NJ 07030-5774  
[www.wiley.com](http://www.wiley.com)

Copyright © 2018 por John Wiley & Sons, Inc.

Nenhuma parte desta publicação pode ser reproduzida, armazenada num sistema de recuperação ou transmitido em qualquer forma ou por qualquer meio, electrónico, mecânico, de gravação, de varrimento ou de outra forma, excepto quando permitido sob as secções 107 ou 108 do 1976 Copyright dos Estados Unidos ato, sem a prévia autorização por escrito do editor. Pedidos para o Publisher para a permissão deve ser dirigida ao Departamento de Permissões, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ

07030, (201) 748-6011, fax (201) 748-6008, ou online em <http://www.wiley.com/go/permissions>.

**marcas:** Wiley, For Dummies, o logotipo Dummies Man, The Way Dummies, Dummies.com, tornando tudo mais fácil, e vestido de comércio relacionado são marcas comerciais ou marcas registradas da John Wiley & Sons, Inc. e / ou suas afiliadas nos Estados Unidos e outros países, e não podem ser utilizados sem autorização escrita. IBM eo logotipo da IBM são marcas registradas da International Business Machines Corporation. Todas as outras marcas comerciais são propriedade dos seus respectivos proprietários. John Wiley & Sons, Inc., não está associada a qualquer produto ou fornecedor mencionado neste livro.

LIMITE DE RESPONSABILIDADE / ISENÇÃO DE GARANTIA: o editor eo AUTOR NÃO FAZEM DECLARAÇÕES OU GARANTIA COM RELAÇÃO À PRECISÃO OU INTEGRALIDADE DO CONTEÚDO DESTE TRABALHO E ESPECIFICAMENTE REJEITAM TODAS AS GARANTIAS, INCLUINDO, SEM LIMITAÇÃO, GARANTIAS DE ADEQUAÇÃO A UM DETERMINADO FIM. NENHUMA GARANTIA podem ser criados ou estendida por vendas ou materiais promocionais. Os conselhos e estratégias contidas AQUI Pode não ser adequado para cada situação. Este trabalho é vendido com o entendimento de que a editora não está envolvida na prestação LEGAL, contabilidade ou outros serviços profissionais. Se a assistência profissional é necessário, os serviços de uma pessoa profissional competente deve ser procurado. Nem o Editor nem o autor SERÁ RESPONSÁVEL POR DANOS dele decorrentes. O fato de uma organização ou WEBSITE é referido neste trabalho como uma citação E / OU UMA FONTE POTENCIAL DE INFORMAÇÕES NÃO SIGNIFICA QUE O AUTOR OU EDITOR APROVA AS INFORMAÇÕES A organização ou site pode fornecer ou recomendações que possa formular. ALÉM DISSO, os leitores devem estar cientes que os sites Internet listados NESTE TRABALHO pode ter mudado ou desapareceram ENTRE quando esta obra foi escrito e quando é lido.

Para informações gerais sobre nossos outros produtos e serviços, ou como criar um personalizado *Para Leigos* livro para o seu negócio ou organização, entre em contato com o nosso Departamento de Desenvolvimento de Negócios nos EUA em **877-409-4177**, contato [info@dummies.biz](mailto:info@dummies.biz) , ou visita [www.wiley.com/go/custompub](http://www.wiley.com/go/custompub).

Para obter informações sobre o licenciamento do *Para Leigos* marca para produtos ou serviços, entre em contato BrandedRights & Licenses@Wiley.com.

ISBN: 978-1-119-45495-3 (PBK); ISBN: 978-1-119-45494-6 (EBK) Fabricado nos Estados Unidos da

América 10 9 8 7 6 5 4 3 2 1

## Agradecimentos do Publisher

Algumas das pessoas que ajudou a trazer este livro para o mercado incluem o  
Segue:

**Projeto editor:** Carrie A. Burchfield

**Gerente Editorial:** Rev Mengle

**Aquisições editor:** Steve Hayes

**Desenvolvimento de negócios**

**Representante:** Bêncão Sue

**IBM Contribuintes:**

Jean-Francois Puget, Nancy Hensley, Brad  
Murphy, Troy Hernandez

# Índice

<b>INTRODUÇÃO .....</b>	<b>1</b>
Sobre este livro .....	1
Pressupostos tolas .....	2
Ícones utilizados neste livro .....	2
 <b>CAPÍTULO 1: Aprendizado de Máquina compreensão .....</b>	<b>3</b>
O que é Machine Learning? .....	4
aprendizagem iterativa a partir de dados .....	5
O que é velho é novo outra vez .....	5
Definindo Big Data .....	6
Big Data em contexto com Machine Learning .....	7
A necessidade de compreender e confiar seus dados .....	8
A importância da Nuvem Híbrida .....	9
Aproveitando o poder de Machine Learning .....	9
análise descritiva .....	10
Análise preditiva .....	10
Os papéis de Estatística e Data Mining com Machine Learning .....	11
Colocando Machine Learning em Contexto .....	12
Approaches to Machine Learning .....	14
aprendizagem supervisionada .....	15
aprendizagem não supervisionada .....	15
aprendizado por reforço .....	16
Redes neurais e aprendizagem profunda .....	17
 <b>CAPÍTULO 2: Aplicando Aprendizagem Automática .....</b>	<b>19</b>
Começando com uma estratégia .....	19
Usando a aprendizagem de máquina para remover preconceitos de estratégia .....	20
Mais dados torna o planejamento mais preciso .....	22
Compreender técnicas de Aprendizado de Máquina .....	22
Amarar métodos de aprendizado de máquina para os resultados .....	23
Aplicando Machine Learning to Business Necessidades .....	23
Compreender por que os clientes estão deixando .....	24
Reconhecendo que tenha cometido um crime .....	25
Prevenir a ocorrência de acidentes .....	26

## **CAPÍTULO 3: Olhando para dentro Machine Learning ..... 27**

O Impacto da Machine Learning sobre Aplicações .....	28
O papel dos algoritmos .....	28
Tipos de algoritmos de aprendizado de máquina .....	29
sistemas de aprendizagem máquina do treinamento .....	33
Preparação de dados .....	34
Identificar dados relevantes .....	34
Governar dados .....	36
O Ciclo de Aprendizagem Máquina .....	37

## **CAPÍTULO 4: Introdução ao Aprendizado de Máquina ..... 39**

Entendendo como Machine Learning pode ajudar .....	39
Concentre-se sobre o problema de negócios .....	40
Trazendo silos de dados em conjunto .....	41
Evitar problemas antes que aconteça .....	42
Obtendo focada no cliente .....	43
Machine Learning requer a colaboração .....	43
A execução de um Projeto Piloto .....	44
Passo 1: Definir uma oportunidade para o crescimento .....	44
Passo 2: Realização de um projeto piloto .....	44
Passo 3: Avaliação .....	45
Passo 4: ações Próxima .....	45
Determinando o melhor modelo de aprendizagem .....	46
Ferramentas para determinar a seleção algoritmo .....	46
Aproximando-seleção da ferramenta .....	47

## **CAPÍTULO 5: Aprendizado de máquina Skills ..... 49**

Definindo as habilidades que você precisa .....	49
Primeiros Educado .....	53
Recursos IBM-Recomendado .....	56

## **CAPÍTULO 6: Usando Machine Learning para fornecer**

<b>Soluções para problemas de negócios .....</b>	<b>57 Aplicando</b>
Machine Learning à saúde do paciente .....	57
Aproveitando a Internet das coisas para criar resultados mais previsíveis .....	58
Proativamente Respondendo a questões que .....	59
Proteção contra Fraudes .....	60

## **CAPÍTULO 7: Dez previsões sobre o futuro**

<b>de Aprendizado de Máquina .....</b>	<b>63</b>
--	-----------

# Introdução

**M** software é projetado de modo que possa manter o ritmo com a mudança nos negócios, e a aprendizagem de máquina está tendo um impacto dramático sobre a maneira de conduzir regras de negócio e lógica. Como isso é diferente? Com os modelos de desenvolvimento de software tradicionais, os programadores escrevem lógica com base no estado atual da empresa e, em seguida, acrescentam dados relevantes. No entanto, a mudança nos negócios tornou-se a norma. É praticamente impossível prever o que muda e transformará um mercado.

O valor da aprendizagem de máquina é que ele permite que você aprenda continuamente a partir de dados e prever o futuro. Este poderoso conjunto de algoritmos e modelos estão sendo usados em todos os setores para melhorar os processos e obter insights sobre padrões e anomalias dentro dos dados.

Mas a aprendizagem de máquina não é um esforço solitário; é um processo de equipe que requer cientistas de dados, engenheiros de dados, analistas de negócios e líderes empresariais para colaborar. O poder da aprendizagem de máquina requer uma colaboração para que o foco esteja em resolver problemas de negócios.

## Sobre este livro

*Machine Learning For Dummies*, IBM Edição Limitada, dá-lhe insights sobre o que a aprendizagem de máquina é sobre tudo e como ele pode afetar a forma como você pode armar dados para obter insights inimagináveis. Seus dados só são tão bons quanto o que você faz com eles e como você gerenciá-los. Neste livro, você descobre tipos de técnicas de aprendizado de máquina, modelos e algoritmos que podem ajudar a alcançar resultados para a sua empresa. Esta informação ajuda tanto para negócios e líderes técnicos aprender como aplicar a aprendizagem de máquina para antecipar e prever o futuro.

# Pressupostos tolas

As informações contidas neste livro é útil para muitas pessoas, mas temos que admitir que nós fizemos algumas suposições sobre quem nós pensamos que você é:

- » Você já está familiarizado com a forma como a aprendizagem de máquina algoritmos estão sendo usados dentro de sua organização para criar um novo software. Você precisa estar preparado para levar sua equipe na direção certa para que a empresa ganha valor máximo do uso desses algoritmos e modelos poderosos.
- » Você está planejando uma estratégia de longo prazo para criar um software que pode resistir ao teste do tempo. Gestão quer ser capaz de aproveitar todos os dados importantes sobre os clientes, colaboradores, clientes potenciais e tendências de negócios. Seu objetivo é estar preparado para o futuro.
- » Você entende o enorme valor potencial dos dados que existe em toda a organização.
- » Você entende os benefícios da aprendizagem de máquina e sua impacto sobre a empresa, e você quer ter certeza de que sua equipe está pronta para aplicar esse poder para se manterem competitivas como novos modelos de negócios surgem.
- » Você é um líder de negócio que quer aplicar o mais importante tecnologias emergentes para ser tão criativo e inovador quanto possível.

## Ícones utilizados neste livro

Os seguintes ícones são usados para destacar informações importantes ao longo do livro:



TIP

Dicas ajudar a identificar informações que precisa de atenção especial.



WARNING

Estes ícones apontar conteúdo que você deve prestar atenção. Destacamos armadilhas comuns em tirar proveito de modelos de aprendizado de máquina e algoritmos.



REMEMBER

Este ícone destaca informações importantes que você deve se lembrar.

## NESTE CAPÍTULO

- » Definir aprendizado de máquina e dados de grande
- » Confiando em seu dados
- » Olhando por isso que a nuvem híbrida é importante
- » Usando aprendizado de máquina e artificial inteligência
- » Compreender as abordagens para aprendizado de máquina

# Capítulo 1

## Aprendizado de Máquina compreensão

**M**achine learning, inteligência artificial (AI) e computing estão dominando conversas sobre como emergentes análises avançadas de dados, aprendizado de máquina, inteligência artificial (AI) e cognitiva para a empresa. Não há debate que os líderes de negócios existentes estão enfrentando novos e inesperados concorrentes. Estas empresas estão olhando para novas estratégias que podem prepará-los para o futuro. Enquanto uma empresa pode tentar estratégias diferentes, todos eles voltam para uma verdade fundamental - você tem que seguir os dados. Neste capítulo, aprofundar o que o valor de aprendizagem de máquina pode ser a sua estratégia de negócios. Como você deve pensar sobre a aprendizagem de máquina? O que você pode oferecer o negócio baseado na técnica de análise avançada que pode ser um divisor de águas?



# O que é Machine Learning?

aprendizado de máquina tornou-se um dos temas mais importantes dentro das organizações de desenvolvimento que estão à procura de formas inovadoras para alavancar os ativos de dados para ajudar a empresa a ganhar um novo nível de compreensão. Por que acrescentar aprendizado de máquina na mistura? Com os modelos de aprendizado de máquina apropriadas, as organizações têm a capacidade de prever continuamente mudanças no negócio de modo que eles são mais capazes de prever o que vem a seguir. Como os dados são constantemente adicionadas, os modelos de aprendizado de máquina garantir que a solução é constantemente atualizado. O valor é simples: Se você usar as fontes de dados mais adequados e em constante mudança no contexto de aprendizagem de máquina, você tem a oportunidade de prever o futuro.

aprendizagem de máquina é uma forma de AI que permite um sistema para aprender a partir de dados em vez de através de uma programação explícita. No entanto, a aprendizagem de máquina não é um processo simples.



**REMEMBER**

aprendizado de máquina usa uma variedade de algoritmos que iteratively aprender com dados para melhorar, descrever os dados e prever resultados. Como os algoritmos ingerir dados de treinamento, é então possível produzir modelos mais precisos com base nesses dados. Um modelo de aprendizagem de máquina é a saída gerada quando você treinar o seu algoritmo de aprendizagem de máquina com dados. Após o treinamento, quando você fornece um modelo com uma entrada, você será dado uma saída. Por exemplo, um algoritmo preditivo irá criar um modelo preditivo. Então, quando você fornecer o modelo de previsão com dados, você receberá uma previsão com base nos dados que treinou o modelo. aprendizagem de máquina é agora essencial para a criação de modelos de análise.

Você provavelmente interagem com aplicativos de aprendizado de máquina sem perceber. Por exemplo, quando você visita um site de e-commerce e começar a ver os produtos e ler comentários, você provavelmente apresentados com outros produtos similares que você pode achar interessante. Estas recomendações não são codificado por um exército de desenvolvedores. As sugestões são servidos para o site através de um modelo de aprendizagem de máquina. O modelo ingere seu histórico de navegação, juntamente com a navegação e aquisição de dados, a fim de apresentar outros produtos similares que você pode querer comprar de outros compradores.

## aprendizagem iterativa a partir de dados

aprendizado de máquina permite modelos para treinar em conjuntos de dados antes de ser implantado. Alguns modelos de aprendizado de máquina são *conectados* e continuamente se adaptam como novos dados são ingeridos. Por outro lado, outros modelos, chamados *modelos de aprendizagem de máquina off-line*, são derivados de algoritmos de aprendizagem de máquina, mas, uma vez implantado, não mudam. Este processo iterativo de modelos on-line leva a uma melhoria nos tipos de associações que são feitas entre elementos de dados. Devido à sua complexidade e tamanho, esses padrões e associações poderia facilmente ter sido esquecido por observação humana. Depois de um modelo foi treinado, estes modelos podem ser usados em tempo real para aprender a partir de dados.



TIP

Além disso, algoritmos complexos pode ser ajustado automaticamente com base em mudanças rápidas de variáveis, tais como dados do sensor, tempo, dados meteorológicos e métricas sentimento do cliente. Por exemplo, inferências podem ser feitas a partir de um modelo de aprendizagem de máquina - Se o tempo muda rapidamente, um modelo de clima previsão pode prever um tornado, e uma sirene de alerta pode ser disparado. As melhorias na precisão são um resultado do processo de formação e de automação que faz parte do aprendizado de máquina. Online algoritmos de aprendizado de máquina refinar continuamente os modelos processando continuamente novos dados em tempo quase real e treinando o sistema para se adaptar à mudança de padrões e associações nos dados.

## O que é velho é novo outra vez

AI e algoritmos de aprendizado de máquina não são novas. O campo da AI remonta à década de 1950. Arthur Lee Samuels, um pesquisador da IBM, desenvolveu um dos programas de aprendizado de máquina mais antigos - um programa de auto-aprendizagem para jogar damas. Na verdade, ele cunhou o termo *aprendizado de máquina*. Sua abordagem de aprendizagem de máquina foi explicada em um artigo publicado no *IBM Jornal de Pesquisa e Desenvolvimento* em 1959.

Ao longo das décadas, técnicas de IA têm sido amplamente utilizado como um método de melhorar o desempenho de código subjacente. Nos últimos anos, com o foco em modelos de computação distribuída e computação mais barato e de armazenamento, tem havido uma onda de interesse em AI e aprendizado de máquina que levou a uma enorme quantidade de dinheiro sendo investido em empresas de software de inicialização. Hoje nós

estão vendo grandes avanços e soluções comerciais. Por que o mercado se tornar real? Há seis elementos fundamentais:

- » Os processadores modernos tornaram-se cada vez mais poderoso e cada vez mais densa. A densidade em relação ao desempenho melhorou drasticamente.
- » O custo de armazenamento e gerenciamento de grandes quantidades de dados tem sido drasticamente reduzida. Além disso, novas inovações de armazenamento levaram a um desempenho mais rápido e a capacidade de analisar conjuntos de dados muito maiores.
- » A capacidade para distribuir o processamento de computação através de aglomerados de computadores melhorou dramaticamente a capacidade de analisar dados complexos em tempo recorde.
- » Há mais conjuntos de dados comerciais disponíveis para apoio analytics, incluindo dados meteorológicos, dados de mídia social, e conjuntos de dados médicos. Muitos destes estão disponíveis como serviços em nuvem e Interfaces de Programação de Aplicações bem definidos (APIs).
- » algoritmos de aprendizado de máquina foram disponibilizados através de comunidades de código aberto com grandes bases de usuários. Portanto, há mais recursos, estruturas e bibliotecas que fizeram o desenvolvimento mais fácil.
- » Visualization ficou mais consumível. Você não precisa ser um cientista de dados para interpretar os resultados, fazendo uso de máquina de aprendizagem mais ampla dentro de muitas indústrias.

## Definindo Big Data

*Big data* é qualquer tipo de fonte de dados que tem pelo menos uma das quatro características compartilhadas, chamadas as quatro Vs:

- » Extremamente grande *Volumes* De dados
- » A capacidade para se mover de que os dados a uma alta *Velocidade* de velocidade
- » Uma constante expansão *Variedade* de fontes de dados
- » *Veracidade* de modo que as fontes de dados verdadeiramente representar a verdade

A precisão de um modelo de aprendizagem de máquina pode aumentar substancialmente se ele é treinado em dados grandes. Sem dados suficientes, você está

tentando tomar decisões em pequenos subconjuntos de seus dados que podem levar à má interpretação de uma tendência ou faltando um padrão que está apenas começando a emergir. Enquanto grandes de dados pode ser muito útil para modelos de aprendizagem máquina do treinamento, as organizações podem usar a máquina de aprendizagem com apenas alguns milhares de pontos de dados.



**WARNING**

Não subestime a tarefa em mãos. Os dados devem ser capazes de ser verificada com base na precisão e contexto. Uma empresa inovadora em um mercado em rápida mudança vai querer implantar um modelo que pode fazer inferências em milissegundos para avaliar rapidamente a melhor oferta para um cliente em situação de risco para mantê-la feliz. É necessário identificar a quantidade certa e tipos de dados que podem ser analisados para impactar os resultados de negócios. Big data incorpora todos os dados, incluindo dados estruturados, não estruturados e semi-estruturados de e-mail, mídias sociais, fluxos de texto, imagens e sensores da máquina.



**WARNING**

Os produtos tradicionais de Business Intelligence (BI) não foram realmente concebido para lidar com as complexidades da constante mudança fontes de dados. ferramentas de BI são normalmente concebidos para trabalhar com, dados bem compreendidos altamente estruturadas, muitas vezes armazenados em um repositório de dados relacional. Estas ferramentas de BI tradicionais normalmente só analisar instantâneos de dados em vez de todo o conjunto de dados. Analytics sobre big data requer tecnologia projetada para reunir, armazenar, gerenciar e manipular grandes quantidades de dados na velocidade certa e no momento certo para ganhar os insights certos. Com a evolução da tecnologia de computação e o surgimento de arquiteturas de nuvem híbrida, agora é possível administrar imensos volumes de dados que anteriormente só poderia ter sido tratado por supercomputadores em grande despesa.

## Big Data em contexto com

## Aprendizado de Máquina

aprendizado de máquina requer o direito conjunto de dados que pode ser aplicado a um processo de aprendizagem. Uma organização não tem que ter grandes dados, a fim de usar técnicas de aprendizado de máquina; no entanto, grandes dados podem ajudar a melhorar a precisão dos modelos de aprendizagem de máquina. Com dados grandes, agora é possível virtualizar dados para que possa ser armazenada de forma mais eficiente e rentável se onpremises ou na nuvem. Além disso, as melhorias na velocidade da rede e fiabilidade ter removido outras limitações físicas de

ser capaz de gerenciar grandes quantidades de dados a uma velocidade aceitável. Adicione a isso o impacto de mudanças no preço e sofisticação da memória do computador, e com todas essas transições de tecnologia, agora é possível imaginar como as empresas podem aproveitar os dados de uma maneira que teria sido inconcebível há apenas cinco anos.



**REMEMBER**

Sem transição da tecnologia acontece em isolamento; mudança acontece quando há um problema de negócios sem solução combinada com a maturação da tecnologia. Há inúmeros exemplos de tecnologias importantes que amadureceram o suficiente para suportar o renascimento da aprendizagem de máquina. Estes vencimento tecnologias de dados grandes incluem a virtualização de dados, processamento paralelo, sistemas de arquivos distribuídos, bancos de dados em memória, da contentorização e micro-serviços. Esta combinação de avanços tecnológicos podem ajudar as organizações a abordar problemas de negócios significativos. As empresas nunca faltava grandes quantidades de dados. Líderes têm sido frustrados por décadas sobre a sua incapacidade de usar a riqueza de fontes de dados para obter insights de seus dados.



**REMEMBER**

Armado com tecnologias de dados grandes e modelos de aprendizagem de máquina, as organizações são capazes de antecipar o futuro e estar melhor preparado para interrupções.

## A necessidade de compreender e confiar seus dados

Não é suficiente simplesmente ingerir grandes quantidades de dados. Fornecendo modelos de aprendizagem de máquina precisos requer que os dados de origem sejam precisas e significativas. Além disso, essas fontes de dados são significativos quando combinados uns com os outros para que o modelo é precisa e confiável. Você tem que entender a origem de suas fontes de dados e se elas fazem sentido quando são combinados.

Além de confiar seus dados, é também importante realizar a limpeza de dados ou arrumação. Limpeza de dados significa que você transformar seus dados em um formulário que pode ser entendido por um algoritmo de aprendizagem de máquina. Por exemplo, os algoritmos de usar números, mas os dados são muitas vezes na forma de palavras. Você tem que transformar essas palavras em números. Além disso, você tem que ter certeza esses números são

sensatamente derivados e internamente consistente. Você precisa decidir como você lida com falta de dados e outras irregularidades de dados.



**REMEMBER**

refinamento de dados fornece a base para a construção de modelos analíticos que fornecem resultados que você pode confiar. O processo de refinamento de dados vai ajudar a garantir que seus dados é oportuna, limpo e bem compreendida.

## A importância da Nuvem Híbrida

Ao se aproximar de aprendizado de máquina e dados grande, muitas organizações descobriram que uma combinação de serviços de nuvem pública e privada é a forma mais pragmática para garantir escalabilidade, segurança e conformidade. Para aprofundar a aprendizagem, uma empresa pode, por exemplo, querem aproveitar Graphics Processing Units (GPUs) na nuvem em vez de construir o seu próprio ambiente baseado em GPU. Esta é uma abordagem híbrida.



**REMEMBER**

A nuvem híbrida é uma combinação de no local e serviços de nuvem pública destinados a trabalhar em uníssono. O ambiente híbrido oferece às empresas a flexibilidade para selecionar o serviço mais adequado para cargas de trabalho específicas com base em fatores críticos, tais como custo, segurança e desempenho.

A computação em nuvem permite que as empresas para testar novos empreendimentos sem os grandes custos iniciais de hardware no local. Ao invés de ir através da aquisição e integração, as equipes podem começar imediatamente a trabalhar com técnicas de aprendizado de máquina. À medida que a organização amadurece, ela pode optar por trazer algum do hardware no local por causa de segurança e controle ou os custos de computação em nuvem que podem aumentar rapidamente.

## Aproveitando o Poder do Aprendizado de Máquina

O papel de análise em processos operacionais de uma organização mudou significativamente ao longo dos últimos 30 anos. As empresas estão experimentando uma progressão nos níveis de análise de vencimento que vão desde análises descritivas para análise preditiva para aprendizado de máquina e computação cognitiva. As empresas têm sido bem sucedidas em

usando a análise para entender tanto de onde eles vieram e como eles podem aprender com o passado para prever o futuro. Eles são capazes de descrever como várias ações e eventos terá impacto sobre os resultados. Enquanto o conhecimento a partir desta análise pode ser usado para fazer previsões, normalmente essas previsões são feitas através de uma lente de expectativas preconcebidas.



**REMEMBER**

cientistas de dados e analistas de negócios têm sido obrigados a fazer previsões baseadas em modelos analíticos que são baseados em dados históricos. No entanto, há sempre fatores desconhecidos que podem ter um impacto significativo sobre os resultados futuros. As empresas precisam de uma forma de construir modelos preditivos que podem reagir e mudar quando há mudanças no ambiente de negócios.

Nesta seção, nós damos-lhe dois tipos de abordagens para análise avançada.

## **análise descritiva**

análise descritiva ajuda os analistas a compreender a realidade atual no negócio. Você precisa entender o contexto para os dados históricos, a fim de compreender a realidade atual de onde a empresa é hoje. Essa abordagem ajuda a organização responder a perguntas como quais estilos produtos estão vendendo melhor neste trimestre, em comparação ao trimestre anterior, e quais regiões estão exibindo o maior para o menor crescimento /.

## **Análise preditiva**

A análise preditiva ajuda a antecipar as mudanças com base na compreensão dos padrões e anomalias dentro desse dados. Com este modelo, o analista assimila uma série de fontes de dados relacionados, a fim de prever os resultados. A análise preditiva utiliza sofisticados algoritmos de aprendizado de máquina para obter insights em curso.



**REMEMBER**

Uma ferramenta de análise preditiva exige que o modelo é constantemente fornecido com novos dados que refletem a mudança nos negócios. Esta abordagem melhora a capacidade da empresa para antecipar as mudanças sutis nas preferências dos clientes, a erosão dos preços, mudanças de mercado e outros fatores que impactarão o futuro dos resultados de negócios.

Com um modelo preditivo, você olhar para o futuro. Por exemplo, você pode responder os seguintes tipos de perguntas:

- » Como pode a experiência web ser transformado para atrair um cliente a comprar com frequência?
- » Como você prever como uma bolsa ou uma carteira irá realizar com base em notícias internacionais e fatores financeiros internos?
- » Que combinação de drogas irá fornecer o melhor resultado para este doente de cancro com base nas características específicas do tumor e sequência genética?

## Os papéis de Estatística e Data Mining com a aprendizagem de máquina

As disciplinas de estatística, mineração de dados, e máquina de aprendizagem, todos têm um papel na compreensão dos dados, descrevendo as características de um conjunto de dados e encontrar relações e padrões em que os dados para construir um modelo. Há uma grande quantidade de sobreposição na forma como as técnicas e ferramentas dessas disciplinas são aplicadas para resolver problemas de negócios.



**REMEMBER**

Muitos dos algoritmos de mineração de dados e de aprendizado de máquina amplamente utilizados estão enraizados na análise estatística clássica. Cientistas de dados combinar fundos da tecnologia com experiência em estatísticas, mineração de dados e aprendizagem de máquina para usar todas as disciplinas em colaboração. Independentemente da combinação de capacidades e da tecnologia utilizados para prever resultados, tendo uma compreensão do problema de negócio, objetivos de negócio e experiência sobre o assunto é fundamental. Você não pode esperar para obter bons resultados, concentrando-se nas estatísticas sozinho, sem considerar o lado do negócio.

Os seguintes pontos destacar como esses recursos se relacionam entre si:

- » **Estatísticas é a ciência da análise dos dados.** clássica ou estatísticas convencionais é inferencial na natureza, o que significa que é usado para chegar a conclusões sobre os dados (vários parâmetros). modelagem estatística é focado principalmente em fazer inferências e compreender as características das variáveis. Máquina de modelos de aprendizagem alavancagem algoritmos estatísticos e aplicá-los a prever analytics. Em um modelo estatístico, uma hipótese é uma forma testável para confirmar a validade do algoritmo específico.





REMEMBER

### » Mineração de dados, que se baseia nos princípios de estatísticas, é

o processo de explorar e analisar grandes quantidades de dados para descobrir padrões em que os dados. Algoritmos são usados para encontrar relações e padrões nos dados, e depois esta informação sobre os padrões é usado para fazer previsões e predições. A mineração de dados é usado para resolver uma série de problemas de negócios, tais como detecção de fraudes, análise de cesta de mercado e análise de rotatividade de clientes. Tradicionalmente, as organizações usam ferramentas de mineração de dados em grandes volumes de dados estruturados, tais como bancos de dados de gestão de relacionamento com clientes ou estoques de peças de aeronaves. O objetivo da mineração de dados é explicar e entender os dados. mineração de dados não se destina a fazer previsões ou fazer backup de hipóteses. Alguns fornecedores de análise fornecer soluções de software que permitem a mineração de uma combinação de dados estruturados e não estruturados de dados. Geralmente, o objectivo da extracção de dados é para extrair os dados a partir de um maior conjunto de dados para efeitos de classificação ou de predição. Na exploração de dados, os dados são agrupados em grupos. Por exemplo, um comerciante pode estar interessado nas características das pessoas que responderam a uma oferta promocional versus aqueles que não responderam à promoção. Neste exemplo, extracção de dados seria usado para extrair os dados de acordo com as duas classes diferentes e analisar as características de cada classe. Um comerciante pode estar interessado em prever quem irá responder a uma promoção. ferramentas de mineração de dados se destinam a apoiar o processo de tomada de decisão humana. Portanto, a mineração de dados se destina a mostrar padrões que podem ser usados por seres humanos. Em contraste, a aprendizagem de máquina automatiza o processo de identificação de padrões que são usados para fazer previsões.

algoritmos de aprendizado de máquina são abordados na próxima seção, "Colocar Machine Learning em Contexto", em maior detalhe, devido à importância desta disciplina para análises avançadas.

## Colocando Machine Learning em Contexto

Para compreender o papel da aprendizagem de máquina, é preciso lhe dar algum contexto. AI, aprendizado de máquina e aprendizagem profunda são todos os termos que são frequentemente mencionados quando se fala de grandes dados, análises e tecnologia avançada. AI pode ser entendido como o

forma mais ampla de sistemas que podem descrever “pensar”. Por exemplo, termostatos que aprendem a sua preferência ou aplicações que podem identificar pessoas e que eles estão fazendo nas fotos pode ser pensado como sistemas de inteligência artificial.

Tal como ilustrado na figura 1/1, existem quatro principais subconjuntos de AI. Neste livro, vamos nos concentrar no aprendizado de máquina. No entanto, a fim de compreender a aprendizagem de máquina, é importante para colocá-lo em perspectiva.

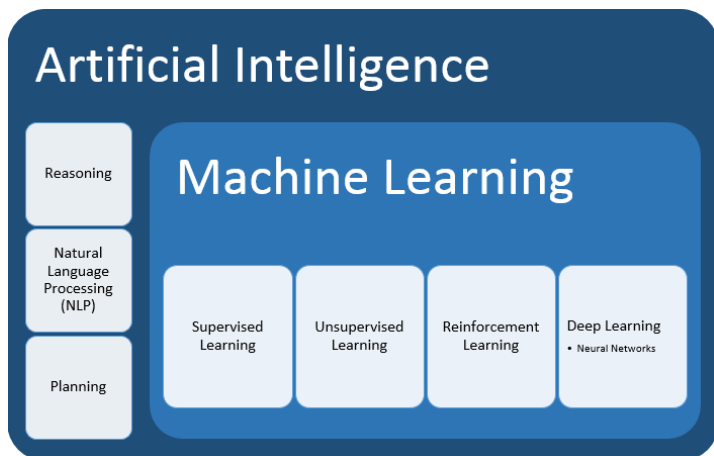


Figura 1-1: AI é a categoria geral que inclui aprendizado de máquina e processamento de linguagem natural.



#### REMEMBER

Quando exploramos aprendizagem de máquina, vamos nos concentrar na capacidade de aprender e adaptar um modelo com base nos dados em vez de programação explícita. No Capítulo 6, vamos nos concentrar na aplicação de aprendizagem de máquina para resolver problemas de negócios.

Antes de nos aprofundarmos os tipos de aprendizagem de máquina, é importante entender os outros subgrupos de AI:

#### » **Raciocínio:** raciocínio máquina permite um sistema para fazer

inferências baseadas em dados. Em essência, o raciocínio ajuda a preencher os espaços em branco quando não há dados incompletos. raciocínio máquina ajuda a dar sentido de dados conectados. Por exemplo, se um sistema tiver dados suficientes e é perguntado “O que é uma temperatura interna segura para comer uma coxa de galinha?”, O sistema seria capaz de lhe dizer que a resposta é 165 graus. o

cadeia lógica seria como se segue: uma perna que é comido (em oposição a uma parte de um instrumento musical) refere-se a uma coxa de galinha, um pé de galinha contém carne de frango escuro, carne de frango escuro precisa de ser cozido a 165 graus, por conseguinte, o resposta é 165 graus. **Nota:** Neste exemplo, o sistema nunca foi treinado explicitamente sobre a temperatura interna segura de coxinhas de frango. Em vez disso o sistema utilizado o conhecimento que tinha de preencher as lacunas de dados.

» **Processamento de Língua Natural (PLN):** PLN é a capacidade de

computadores de trem para compreender tanto texto escrito e fala humana. técnicas de PLN são necessários para capturar o significado do texto não estruturado a partir de documentos ou comunicação com o usuário. Portanto, a PLN é a principal maneira que os sistemas podem interpretar texto e língua falada. PLN é também uma das tecnologias fundamentais que permite que pessoas não-técnicas para interagir com tecnologias avançadas. Por exemplo, em vez de precisar de código, a PLN pode ajudar os usuários a pedir um sistema de perguntas sobre conjuntos de dados complexos. Ao contrário de informações de banco de dados estruturado que se baseia em esquemas para adicionar contexto e significado aos dados, informações não estruturadas deve ser analisado e marcou para encontrar o significado do texto. Ferramentas necessárias para PLN incluem categorização, ontologias, tapping, catálogos, dicionários e modelos de linguagem.

» **Planejamento:** planejamento automatizado é a capacidade de um inteligente

sistema para agir de forma autónoma e flexível para a construção de uma sequência de acções para alcançar um objectivo final. Em vez de um processo de tomada de decisão de pré-programado que vai de A para B para C, para chegar a um produto final, o planeamento automatizado é complexo e requer um sistema para adaptar com base no contexto em torno do desafio dada.

## Abordagens para Aprendizado de Máquina

técnicas de aprendizagem de máquina são necessárias para melhorar a precisão dos modelos preditivos. Dependendo da natureza do problema de negócio a ser abordada, existem diferentes abordagens com base no tipo e volume dos dados. Nesta seção, discutimos as categorias de aprendizagem de máquina.

## aprendizagem supervisionada

aprendizado supervisionado normalmente começa com um conjunto estabelecido de dados e uma certa compreensão de como os dados são classificados. aprendizagem supervisionada é destinada para encontrar padrões em dados que podem ser aplicados a um processo de análise. Estes dados rotulou características que definem o significado dos dados. Por exemplo, pode haver milhões de imagens de animais e incluem uma explicação do que cada animal é e então você pode criar um aplicativo de aprendizado de máquina que distingue um animal de outro. Ao rotular esses dados sobre os tipos de animais, você pode ter centenas de categorias de espécies diferentes. Porque os atributos e o significado dos dados foram identificados, isso é bem compreendido pelos usuários que estão treinando os dados modelados de forma que ele **se encaixa os detalhes das etiquetas. Quando o rótulo é contínua, é uma regressão; *classificação*.** Em essência, a regressão utilizada para aprendizagem supervisionada ajuda a entender a correlação entre as variáveis. Um exemplo de aprendizagem supervisionada é a previsão do tempo. Usando análise de regressão, previsão do tempo leva em conta conhecido padrões climáticos históricos e as condições atuais para fornecer uma previsão sobre o tempo.



### TIP

Os algoritmos são treinados utilizando exemplos pré-processados, e, neste ponto, o desempenho dos algoritmos é avaliada com dados de teste. Ocasionalmente, os padrões que são identificadas em um subconjunto dos dados não pode ser detectado na população maior de dados. Se o modelo está apto a representar apenas os padrões que existem no subconjunto de **treinamento, você cria um problema chamado *overfitting*. Overfitting significa que o seu modelo é** precisamente atento para os dados do treino, mas pode não ser aplicável para grandes conjuntos de dados desconhecidos. Para se proteger contra overfitting, teste precisa ser feito em relação aos dados rotulados imprevistas ou desconhecidos. Usando dados imprevistas para o conjunto de teste pode ajudá-lo a avaliar a precisão do modelo na previsão de resultados e resultados. modelos de formação supervisionados tem ampla aplicabilidade a uma variedade de problemas de negócios, incluindo a detecção de fraudes, soluções de recomendação, reconhecimento de voz, ou análise de risco.

## aprendizagem não supervisionada

aprendizagem não supervisionada é mais adequado quando o problema requer uma enorme quantidade de dados que são unlabeled. Por exemplo, aplicações de mídia social, como Twitter, Instagram, Snapchat, e

assim por diante, todos têm grandes quantidades de dados não marcados. Compreender o significado por trás desses dados requer algoritmos que podem começar a entender o significado com base em ser capaz de classificar os dados com base nos padrões ou clusters que encontra. Portanto, o aprendizado supervisionado realiza um processo iterativo de análise de dados sem intervenção humana. aprendizado não supervisionado é usado com e-mail a tecnologia de detecção de spam. Existem demasiadas variáveis em e-mails legítimos e de spam para um analista de bandeira e-mail em massa não solicitados. Em vez disso, classificadores de aprendizado de máquina com base no agrupamento e de associação são aplicadas, a fim de identificar e-mails indesejados.



REMEMBER

dados de segmento de algoritmos de aprendizagem não supervisionada em grupos de exemplos (clusters) ou grupos de características. Os dados sem rótulo cria os valores dos parâmetros e classificação dos dados. Em essência, este processo adiciona rótulos para os dados de modo que se torna supervisionadas. aprendizagem não supervisionada pode determinar o resultado quando há uma enorme quantidade de dados. Neste caso, o desenvolvedor não conhece o contexto dos dados a serem analisados, de modo a rotulagem não é possível nesta fase. Portanto, a aprendizagem não supervisionada pode ser usado como o primeiro passo antes de passar os dados para um processo de aprendizagem supervisionada.



REMEMBER

algoritmos de aprendizagem sem supervisão pode ajudar as empresas a entender grandes volumes de dados novos, não marcados. De forma semelhante ao aprendizado supervisionado (consulte a seção anterior), esses algoritmos procurar padrões nos dados; no entanto, a diferença é que os dados não estiver compreendida. Por exemplo, na área da saúde, coleta de enormes quantidades de dados sobre uma doença específica pode ajudar os profissionais a obter insights sobre os padrões de sintomas e relacionar aqueles para os resultados de pacientes. Levaria muito tempo para rotular todas as fontes de dados associados com uma doença como a diabetes. Portanto, uma abordagem de aprendizagem não supervisionada pode ajudar a determinar os resultados mais rapidamente do que uma abordagem de aprendizagem supervisionada.

## aprendizado por reforço

aprendizado por reforço é um modelo de aprendizagem comportamental. O algoritmo recebe o feedback a partir da análise dos dados para que o usuário é guiado para o melhor resultado. aprendizado por reforço difere de outros tipos de aprendizado supervisionado porque o sistema não é treinada com o conjunto de dados de amostra. Em vez disso, o sistema aprende por tentativa e erro. Portanto, uma sequência de decisões de sucesso resultará em sendo o processo “reforçada” porque melhor resolve o problema na mão.



#### REMEMBER

Uma das aplicações mais comuns de reforço de aprendizagem está em robótica ou jogar o jogo. Tomemos o exemplo da necessidade de treinar um robô para navegar um conjunto de escadas. O robô muda sua abordagem para navegar no terreno com base no resultado de suas ações. Quando o robô cai, os dados são recalibrados para que os passos são navegados diferente até que o robô é treinado por tentativa e erro para entender como subir escadas. Em outras palavras, o robô aprende baseado em uma sequência bem sucedida de ações. O algoritmo de aprendizagem tem que ser capaz de descobrir uma associação entre a meta de subir escadas com sucesso sem cair e a sequência de eventos que levam ao resultado.

aprendizado por reforço é também o algoritmo que está sendo usado para carros de auto-condução. De muitas maneiras, a formação de um carro de auto-condução é incrivelmente complexa, porque há tantos obstáculos potenciais. Se todos os carros na estrada fosse autônoma, tentativa e erro seria mais fácil de superar. No entanto, no mundo real, os motoristas humanos muitas vezes pode ser imprevisível. Mesmo com este cenário complexo, o algoritmo pode ser otimizada ao longo do tempo para encontrar maneiras de se adaptar ao estado onde as ações são recompensados. Uma das maneiras mais fáceis de pensar sobre a aprendizagem do reforço é a maneira um animal é treinado para tomar medidas com base em recompensas. Se o cão recebe um tratamento cada vez que ele se senta no comando, ele vai tomar esta ação de cada vez.

## Redes neurais e aprendizagem profunda

aprendizagem profundo é um método específico de aprendizagem máquina que incorpora redes neurais em camadas sucessivas, de modo a aprender a partir de dados de uma forma iterativa. aprendizagem profunda é especialmente útil quando você está tentando aprender os padrões de dados não estruturados.

aprendizagem profunda - redes neurais complexos - são projetados para imitar a forma como o cérebro humano funciona para que os computadores podem ser treinados para lidar com abstrações e problemas que são mal definidas. A criança média de cinco anos de idade, pode facilmente reconhecer a diferença entre o rosto do professor e a face do guarda de trânsito. Em contraste, o computador tem que fazer um monte de trabalho para descobrir quem é quem. Redes neurais e aprendizagem profunda são frequentemente utilizados em reconhecimento de imagem, fala e aplicações de visão de computador.

Uma rede neural é constituída por três ou mais camadas: uma camada de entrada, uma ou várias camadas ocultas e uma camada de saída. Os dados são ingeridos através da camada de entrada. Em seguida, os dados são modificados na camada escondida e as camadas de saída com base nos pesos aplicados ao

esses nós. A rede neural típica pode consistir de milhares ou mesmo milhões de nós de processamento simples que são densamente interligadas. A aprendizagem profunda termo é usado quando há várias camadas escondidas dentro de uma rede neural. Usando uma abordagem iterativa, uma rede neural ajusta continuamente e faz inferências até um ponto de parada específico é atingido. As redes neurais são frequentemente utilizados para aplicações de reconhecimento de imagem e visão por computador.

aprendizagem profunda é uma técnica de aprendizado de máquina que utiliza redes neurais hierárquicos para aprender a partir de uma combinação de algoritmos não supervisionados e supervisionados. aprendizagem profunda é muitas vezes chamado de sub-disciplina de aprendizagem de máquina. Normalmente, aprendizagem profunda aprende a partir de dados não marcados e não estruturados. Enquanto o aprendizado profundo é muito semelhante a uma rede neural tradicional, ele terá muitas camadas mais escondidos. Quanto mais complexo o problema, as camadas mais escondido lá será no modelo.



**REMEMBER**

Existem muitas áreas onde a aprendizagem profunda terão um impacto sobre as empresas. Por exemplo, o reconhecimento de voz terá aplicações em tudo, desde automóveis a gestão de clientes. Na Internet of Things (Internet das coisas) aplicações de manufatura, aprendizagem profunda pode ser usado para prever quando uma máquina vai funcionar mal. algoritmos de aprendizagem profundas pode ajudar agentes da lei acompanhar os movimentos de um suspeito conhecido.

- » **Começando com a sua estratégia**
- » Olhando para técnicas de aprendizagem de máquina no problema de negócios
- » Amarrando aprendizagem de máquina para os resultados
- » **Compreender os usos comerciais de**  
aprendizado de máquina

## Capítulo 2

# Aplicando Aprendizagem Automática

**W**os dados gerados por sua empresa para antecipar a mudança de negócios e planos para o futuro. Embora seja claro que o aprendizado de máquina é um conjunto sofisticado de tecnologias, que só é valioso quando você encontrar maneiras de amarrar tecnologia para resultados. Sua empresa não é estático; Portanto, como você aprende mais e mais dos seus dados, você pode estar preparado para a mudança negócio.

## Começando com uma estratégia

Antes de definir a estratégia, você tem que entender o problema que você está tentando resolver. Como as empresas passam por grandes transições de estratégia, alguns desafios se apresentam. Qual é o status de negócios existentes e envolvimento do cliente existente? O que o futuro reserva para o que os clientes vão comprar e esperar de você no futuro? A resposta óbvia é a de perguntar aos clientes se eles estão felizes e que eles vão comprar no futuro. Enquanto este é um ponto som de partida, não é suficiente. Os clientes que estão felizes um minuto se tornam infelizes quando algo transformacional vem. Se você fizer tradicional



Business Intelligence análise (BI), você terá uma boa noção de onde sua empresa tem sido no passado, mas não para onde está indo no futuro.



**REMEMBER**

Sua empresa não é estático; grande parte das nuances e conhecimento sobre seus clientes está escondido dentro de dados estruturados, não estruturados e semi-estruturados. O valor de técnicas de aprendizado de máquina é para ser capaz de descobrir os padrões e anomalias neste enorme quantidade de dados. Selecionar os algoritmos de aprendizagem de máquina certa combinados com as fontes de dados adequadas ajuda a determinar o que está próximo.

## **Usando a aprendizagem de máquina para remover preconceitos de estratégia**

Normalmente, exercícios de planejamento e estratégia estratégico começar por ganhar insights sobre a satisfação do cliente e necessidades futuras. Onde está o mercado indo? Quais são as ameaças competitivas que poderiam afetar a empresa? Mas isto não é o suficiente. Mesmo os melhores consultores de estratégia não pode antecipar o súbito aparecimento de novas descobertas ou novas tendências.



**WARNING**

Uma das armadilhas que a liderança da empresa cai em é seus pressupostos e preconceitos. Demasiadas vezes a gestão da empresa olha para os dados apresentados e interpreta os resultados através de sua própria lente. É o negócio sustentável à luz dos concorrentes emergentes com modelos de negócios imprevistos? Embora seja fácil de ser pego de surpresa da mudança, as sementes da mudança existe. No entanto, esses indicadores principais são muitas vezes enterrado dentro de enormes quantidades de dados não estruturados ou semi-estruturados.

Para ganhar o benefício a partir de uma enorme quantidade de dados não estruturados, é importante para compreender verdadeiramente essas fontes de dados. Qual é a fonte dos dados? Quem tem manipulado os dados? São as fontes de dados confiável? As primeiras experiências em análises avançadas muitas vezes resultaram em resultados decepcionantes porque os analistas agarrou fontes de dados sem habilitação los primeiro. Antes de agir, os dados têm que ser verificado como limpo e preciso. Depois que você está confiante de que você está usando dados precisos para resolver o seu problema de negócio, as abordagens de aprendizado de máquina pode fornecer insights significativos. Ao mesmo tempo, você tem que ter certeza que você tem dados suficientes para descobrir os padrões e anomalias dentro desse dados.



#### TIP

Após a qualidade dos dados é boa, é importante entender o contexto dos dados que está sendo aplicada ao problema. Por exemplo, se uma árvore está a perder suas folhas no meio do verão, é um sinal de que a árvore é saudável. A mesma árvore que perdeu folhas no meio de um dia frio de inverno é uma ocorrência normal. Portanto, sem compreender o contexto dos dados, você provavelmente vai interpretar mal os resultados. Ao mesmo tempo, há uma atenção considerável pago a correlação entre elementos de dados. Quais são as relações entre as condições? No exemplo da saúde das árvores, há uma correlação direta entre as estações e a cor e quantidade de folhas nas árvores. Mas você também tem que ter cuidado com correlações. Você pode encontrar uma correlação que não faz sentido porque o contexto está errado. Não pode parecer ser uma correlação entre as folhas que caem das árvores e o número de casacos de ser comprado online. Embora ambos os eventos estão acontecendo porque o clima é mais frio, não há nenhuma relação entre as árvores e casacos.

Para o negócio de utilizar eficazmente a aprendizagem de máquina para apoiar a estratégia de negócios, você precisa destes métodos estatísticos para encontrar padrões e anomalias nestes conjuntos de dados. Com os melhores dados disponíveis e no volume certo e o melhor nível de limpeza, é possível criar um modelo usando o algoritmo de aprendizado de máquina mais adequado com base no problema de negócio a ser abordado. Este modelo é apenas o começo do fluxo de trabalho de aprendizagem de máquina.

Ao alavancar grandes quantidades de dados, é possível modelar os dados, treinar os dados e, em seguida, começar a aprender a partir desses dados, a fim de melhorar a capacidade de tomar decisões. O valor de aprender a partir de dados significa que o sistema de aprendizagem de máquina é capaz de olhar para os padrões subjacentes e anomalias que não são necessariamente óbvios. Existem relações entre o que os clientes compram com o tempo para reparar? Há impactos do clima sobre as vendas durante um período de tempo? Há indicações nos dados de mídia social que indicam mudanças sutis na percepção do cliente ou padrões de compra? Ser capaz de modelar grandes quantidades de dados de diferentes fontes de dados pode adicionar insights que nenhum humano poderia ter entendido por simplesmente confiar em dados disponíveis de forma isolada.



#### WARNING

Tem havido muita discussão sobre a correlação de dados como um método analítico. Embora a correlação de dados é extremamente importante, às vezes pode ser enganosa. Não pode parecer ser uma correlação entre o consumo de suco de laranja em junho e o aumento da

acidentes de trânsito no mesmo mês, mas não há nenhuma relação causal. Portanto, enquanto a correlação pode ser útil em certos casos, pode também levar a imprecisões. É por isso que o contexto é ainda mais importante. Se houvesse um contexto útil entre suco de laranja e acidentes de trânsito, então a correlação seria útil. Portanto, como você se move para alavancar a aprendizagem de máquina como parte do processo de planejamento e estratégia, você precisa fazer de aprendizado de máquina e análise avançadas ferramentas indispensáveis.

## **Mais dados torna o planejamento mais preciso**

Que diferença poderia fazer a aprendizagem de máquina em estratégia de negócio? Tomemos o exemplo de um negócio que executa uma análise de satisfação do cliente de dados tradicional. Ao analisar os dados, torna-se claro que algumas anomalias nos dados existir. Devido ao conjunto de dados a ser utilizado, o analista joga fora os dados que não estejam em conformidade, assumindo que esses dados não é preciso. No entanto, se existisse mais dados, pode-se claro que essas anomalias que foram assumidos para ser erros são realmente uma indicação de uma mudança nos padrões de compra dos clientes ou a satisfação do cliente. À medida que mais dados são adicionados em um modelo, treinados e analisados com os algoritmos de aprendizado de máquina mais adequadas, torna-se cada vez mais claro que há mudanças que impactarão diretamente o futuro do negócio.

Por exemplo, os cientistas de dados vendo algumas mudanças sutis vai começar a adicionar novas fontes de dados que irão fortalecer ou desmascarar uma análise estatística sobre a mudança negócio ou crescimento. Ao longo do tempo à medida que mais dados são ingeridos para o modelo, o sistema aprende e ganha mais conhecimento e mais sofisticação, a fim de prever o futuro. Portanto, a aprendizagem de máquina torna-se um parceiro de valor inestimável no planejamento estratégico.

## **Compreender técnicas de Aprendizado de Máquina**

A fim de garantir que seus cientistas de dados estão usando as técnicas de aprendizado de máquina certa para alcançar seus objetivos de negócio, é importante entender como a sua organização pode melhor aplicar estas técnicas avançadas para gerir o seu crescimento e manter o foco nas oportunidades emergentes.

aprendizagem de máquina é uma abordagem sistemática para alavancar algoritmos e modelos avançados para treinar continuamente dados e de teste com dados adicionais para começar a aplicar os algoritmos de aprendizado de máquina mais adequado para um problema (discutimos isso com mais detalhes no Capítulo 1). A vantagem de aprendizagem de máquina é que é possível alavancar algoritmos e modelos para prever resultados. O truque é garantir que os cientistas de dados que fazem o trabalho estiver usando os algoritmos corretos, ingerindo os dados mais apropriados (que é preciso e limpo), e utilizando os melhores modelos desempenho. Se todos estes elementos se unem, é possível treinar continuamente o modelo e aprender com os resultados, aprendendo com os dados. A automação desse processo de modelagem, treinamento do modelo e teste leva a previsões precisas para apoiar a mudança nos negócios.

## Amarrar métodos de aprendizado de máquina para os resultados

técnicas de aprendizado de máquina têm o potencial para remodelar mercados inteiros e estratégias de negócios. Por exemplo, as técnicas de aprendizado de máquina estão sendo usados para transformar a indústria automobilística com carros de auto-condução. algoritmos de aprendizado de máquina e modelos estão revolucionando a forma como uma imagem de raios-x é analisada. aprendizado de máquina pode fornecer maneiras proativas de vulnerabilidades de segurança previstas que podem ser reparados antes que o dano está feito. Existem centenas de diferentes soluções que podem ser criados que dependem de técnicas de aprendizado de máquina que podem transformar indústrias inteiras.



REMEMBER

Diferentes abordagens e algoritmos existem para a aprendizagem de máquina, dependendo do problema a ser abordado. Você precisa entender o problema que você está tentando resolver. O modelo que você projetar representará uma compreensão dos dados e sua capacidade de prever resultados com base nesses dados.

## Aplicando Machine Learning às necessidades de negócios

aprendizado de máquina oferece um valor potencial para as empresas que tentam aproveitar os dados grandes e ajuda-los a entender melhor as mudanças sutis no comportamento, preferências, ou a satisfação do cliente.

Os líderes empresariais estão começando a perceber que muitas coisas acontecem dentro de suas organizações e com as suas indústrias que não podem ser compreendidas através de uma consulta. Não são as perguntas que você sabe; é os padrões ocultos e anomalias enterrados na dados que podem ajudar ou prejudicar você. Nesta seção, nós fornecemos alguns exemplos de como as empresas estão começando a usar técnicas de aprendizagem de máquina para criar diferenciação negócio.

## Compreender por que os clientes estão deixando

Você já ouviu falar, “Custa muito menos para manter um cliente existente do que para ganhar um novo cliente”? a rotatividade de clientes é um problema constante em certas indústrias, como telecomunicações, varejo e serviços financeiros.

A compreensão de como impedir os clientes de sair é mais importante do que nunca. Estamos em uma era onde as empresas emergentes estão oferecendo novos modelos de negócios inovadores. Por exemplo, os fornecedores de serviços de telefonia móvel utilizado para exigir um contrato de dois anos, que foi estendido cada vez que o serviço mudou. Como o cenário competitivo mudou, as empresas descobriram que eles tinham que se livrar dos contratos. Esta mudança foi benéfica para os clientes, mas resultou em um aumento enorme na rotatividade de clientes. Sem a proteção de contratos com clientes, empresas de móveis estão se voltando para novas abordagens para manter os clientes.



### REMEMBER

A fim de evitar a rotatividade de clientes, é fundamental que você tenha dados suficientes sobre a história do cliente, suas preferências, os serviços que ele comprou no passado, e suas queixas. Em um mercado altamente estável, essa abordagem para análise poderia ter sido um preditor do futuro. Mas em mercados voláteis, esta abordagem não vai funcionar. Você tem que ser capaz de antecipar as mudanças de mercado e mudanças nos padrões de compra dos clientes. Usando modelos de aprendizagem de máquina pode ajudar a prever as mudanças que irão impactar a receita. Em essência, a operadora de celular precisa ser capaz de olhar para os padrões de dados, bem como anomalias. A operadora de celular tem a vantagem de ter acesso a grandes volumes de dados através de muitos clientes diferentes. Usando o algoritmo certo, o fornecedor pode criar um modelo que mapeia os tipos de ofertas e promoções que irá reter clientes e adicionar novos. Quanto vai custar para reter e adicionar novos clientes? Será que novos planos de reduzir significativamente a receita? Será que os gastos justificam os esforços? Estes

são os tipos de previsões de que uma técnica de aprendizado de máquina pode proporcionar.

Qual é a diferença entre uma abordagem BI tradicional e uma abordagem de aprendizagem de máquina para a rotatividade de clientes? Com o BI tradicional, a organização é capaz de entender o que aconteceu no passado e pode avaliar tendências de fidelidade do cliente. Em contraste, a máquina algoritmo de aprendizagem cria um modelo que traz em quantidades maciças de ambos os dados internos e externos. Após os dados são treinados e testados, os analistas podem começar a antecipar mudanças nas preferências dos clientes. O modelo pode ser capaz de antecipar como padrões de compra dos clientes irá mudar no futuro.



**REMEMBER**

aprendizado de máquina utiliza algoritmos estatísticos como base para a criação de um modelo que pode aprender e prever. Os modelos mais comuns utilizados para modelos preditivos para a análise de churn são algoritmos estatísticos de classificação, tais como regressão logística e redes neurais.

## Reconhecendo que tenha cometido um crime

Os departamentos de polícia têm uma tarefa difícil quando o controle de criminosos. Cada vez mais, há mais e mais câmeras em bairros que ajudam a identificar a atividade ilegal. Mas quem cometeu o ato? Enquanto uma imagem pode valer mil palavras, sem alguém para identificar o mau ator, não é fácil de resolver crimes. Uma das formas de aplicação da lei está tentando alavancar dados de imagem é através do uso de aprendizagem de máquina.



**REMEMBER**

Especificamente, algoritmos de aprendizagem profundas e algoritmos dependentes de redes neurais são os mais adequados para lidar com reconhecimento facial. Em essência, as redes neurais têm a intenção de imitar o cérebro humano. Usando um algoritmo de rede neural, as pessoas podem identificar agrupamentos e padrões em imagens. análise de imagem pode indexar e vídeo de eventos Por classificar objetos em diferentes categorias, tais como pessoas, carros, estradas ou postes. Além disso, os algoritmos de reconhecimento facial pode ser usado para digitalizar seções de uma fotografia de uma pessoa de uma forma que elimina dados irrelevantes que não é útil. Os elementos mais importantes necessários para identificar uma pessoa incluem os olhos, nariz, boca e coisas como cicatrizes. Ao coletar grandes quantidades de dados de imagens faciais, o algoritmo pode identificar padrões nas caras.

rostos. Algumas das técnicas de redes neurais emergentes permitir este tipo de treinamento deve ser feito com poucos dados, o que torna estes sistemas mais prático para uma força policial.

Como é que uma força policial tirar proveito deste tipo de rede neural? A solução incorpora dados de imagem de criminosos conhecidos. Ele inclui dados coletados por câmeras de vigilância, bem como imagens de indivíduos suspeitos que podem estar envolvidos em crimes localmente. Quando um crime acontece, como um assalto a uma loja local, as imagens das câmeras podem identificar os rostos dos indivíduos envolvidos. Estas imagens podem ser comparados com a quantidade de dados. Basicamente, o modelo está olhando para corresponder ao padrão de um rosto específico contra a coleção de imagens para ver se há uma correspondência. Se a polícia pode encontrar o jogo, eles serão capazes de fazer rapidamente uma prisão sem primeiro tomar o tempo para entrevistar testemunhas e passar horas revendo armazenar vídeos.

## Prevenir a ocorrência de acidentes

Muitas indústrias dependem de abordagens de manutenção preventiva sofisticados para garantir que os processos e sistemas são seguros e funcionam como esperado. Indústrias, tais como a fabricação, petróleo e gás, e utilitários de sucesso ou fracasso com base na sua capacidade para evitar acidentes. Embora seja comum ter um cronograma de manutenção, que muitas vezes não é suficiente. Por exemplo, pode haver condições ambientais que impacto as operações de uma máquina ou sistema. Por exemplo, pode haver uma falha de um sistema de aquecimento ou de ar condicionado. Não poderia haver uma mudança dramática em condições meteorológicas que poderiam impactar máquinas.



**TIP**

algoritmos de aprendizado de máquina pode ser aplicado a manutenção preventiva em uma série de maneiras. Por exemplo, um algoritmo de regressão pode ser usada como base para um modelo que pode prever tempo até a falha de uma máquina. Vários algoritmos de classificação pode ser usada para modelar os padrões associados com as falhas da máquina. Os dados gerados pelos sensores fornece um enorme volume de dados semi-estruturado que pode modelar e comparar os padrões de desempenho de modo a que uma anomalia de desempenho normal pode ser detectada.

» Transformando aplicações através

aprendizado de máquina

» Compreender seus dados

» Olhando para o ciclo de aprendizagem de máquina

## Capítulo 3

# Olhando para dentro Machine Learning

**M**áquinas ajudam as organizações a transformar a sua compreensão de dados. Esta abordagem tecnológica é dramaticamente diferente das maneiras em que as organizações tradicionalmente alavancadas dados. Em vez de começar com lógica de negócios e, em seguida, aplicar dados, técnicas de aprendizado de máquina permitem que os dados para criar a lógica. Um dos maiores benefícios desta abordagem é remover hipóteses de negócio e preconceitos que podem causar líderes para adaptar uma estratégia que pode não ser o melhor.

aprendizado de máquina requer um foco na gestão dos dados corretos que é bem preparados. As organizações também devem ser capazes de selecionar os algoritmos adequados que podem fornecer modelos bem desenhados. O trabalho não termina aí. aprendizado de máquina requer um ciclo de gerenciamento de dados, modelagem, treinamento e testes. Neste capítulo, vamos nos concentrar na sustentação tecnologia que suporta soluções de aprendizado de máquina.



# O Impacto da Machine Learning sobre Aplicações

Fizemos uma declaração ousada que, com aprendizagem de máquina você começa com os dados e deixar que os dados levá-lo à lógica. Como é que uma empresa executar no objetivo? Como tudo no desenvolvimento de aplicações complexas e de implantação, que requer um processo de planejamento para a compreensão do problema de negócio que precisa ser resolvido e recolher as fontes de dados corretos.

Como é que esta abordagem para a criação de aplicações têm um impacto sobre o negócio? Quando a construção de aplicações de lógica, você assume que os processos de negócios permanecerá constante. No entanto, a realidade é que os processos de mudança. Se você pode começar por modelagem de dados, ele irá levá-lo a mudanças no processo e lógica. Portanto, a aprendizagem de máquina pode fazer a criação de aplicações muito mais dinâmico e eficaz.

## O papel dos algoritmos

Nenhuma discussão sobre a aprendizagem de máquina seria completa sem uma seção dedicada aos algoritmos.



**REMEMBER**

Os algoritmos são um conjunto de instruções para um computador em como interagir com, manipular e transformar dados. Um algoritmo pode ser tão simples como uma técnica para adicionar uma coluna de números ou tão complexo como identificar o rosto de alguém em uma imagem.

Para fazer um algoritmo operacional, ele deve ser composto como um programa que os computadores podem entender. algoritmos de aprendizado de máquina são mais frequentemente escritos em uma das várias linguagens: Java, Python, ou R. Cada uma dessas línguas incluem bibliotecas de aprendizagem de máquinas que suportam uma variedade de algoritmos de aprendizado de máquina. Além disso, essas línguas têm comunidades de usuários ativos que contribuem regularmente código e discutir ideias, desafios e abordagens para problemas de negócios.



**WARNING**

algoritmos de aprendizado de máquina são diferentes de outros algoritmos. Com a maioria dos algoritmos, um programador começa introduzindo o algoritmo. No entanto, com máquina de aprender o processo é invertido. Com aprendizagem de máquina, os dados em si cria o modelo. Quanto mais dados que é adicionado ao algoritmo, o mais sofisticado algoritmo torna. Como o algoritmo de aprendizado de máquina

é exposto a mais e mais dados, é capaz de criar algoritmo cada vez mais precisas.

## Tipos de algoritmos de aprendizado de máquina

Selecionando o algoritmo certo é parte ciência e arte parte. Dois cientistas de dados encarregado de resolver o mesmo desafio de negócios podem escolher diferentes algoritmos para aproximar-se do mesmo problema. No entanto, compreender as diferentes classes de algoritmos de aprendizagem de máquina ajuda os cientistas de dados identificar os melhores tipos de algoritmos. Esta secção dá-lhe uma breve visão geral dos principais tipos de algoritmos de aprendizagem de máquina.

### Bayesian

*algoritmos Bayesian* permitir que os cientistas de dados para codificar crenças anteriores sobre o que os modelos deve ser semelhante, independente do que os estados de dados. Com tanto foco em dados que definem o modelo, você pode se perguntar por que as pessoas estariam interessados em algoritmos Bayesianos. Esses algoritmos são especialmente úteis quando você não tem grandes quantidades de dados para treinar com confiança um modelo.



REMEMBER

Um algoritmo Bayesian faria sentido, por exemplo, se você tem conhecimento prévio a alguma parte do modelo e pode, pois o código que diretamente. Vamos tomar o caso de um sistema de diagnóstico por imagem médica que procura por distúrbios pulmonares. Se um estudo jornal publicado estima a probabilidade de distúrbios pulmonares diferentes com base no estilo de vida, essas probabilidades podem ser codificados no modelo.

### Clustering

*Clustering* é uma técnica bastante simples de entender - objetos com parâmetros semelhantes são agrupados (em um cluster). Todos os objectos num grupo são mais semelhantes entre si do que os objectos de outros agrupamentos. Clustering é um tipo de aprendizagem não supervisionada, porque os dados não são rotulados. O algoritmo interpreta os parâmetros que compõem cada item e grupos-los em conformidade.

### árvore de decisão

*algoritmos de decisão da árvore* utilizar uma estrutura de ramificação para ilustrar os resultados de uma decisão. árvores de decisão pode ser usado para mapear os possíveis resultados de uma decisão. Cada nó de uma árvore de decisão representa um resultado possível. Percentagens são atribuídos a nós com base na probabilidade do resultado ocorrendo.



#### TIP

As árvores de decisão são por vezes usados para campanhas de marketing. Você pode querer prever o resultado do envio de clientes e prospects um cupão de 20 por cento. Você pode quebrar os clientes em quatro segmentos:

- » Persuadables que provavelmente vai fazer compras, se receber uma extensão
- » que as coisas que vão comprar não importa o que
- » causas perdidas que nunca vai comprar
- » clientes frágeis que podem reagir negativamente a uma divulgação tentativa

Se você enviar uma campanha de marketing, você claramente querem evitar o envio de itens para três dos grupos porque ou não vai responder, loja de qualquer maneira, ou realmente **respondem negativamente**. Segmentação do *persuadables* vai lhe dar o melhor retorno sobre o investimento (ROI). Uma árvore de decisão vai ajudá-lo a mapear estes quatro grupos de clientes e organizar prospects e clientes com base em quem vai reagir melhor à campanha de marketing.

## Redução de dimensionalidade

*Redução de dimensionalidade* ajuda os sistemas de remover dados que não é útil para a análise.

Este grupo de algoritmos é usado para remover dados redundantes, valores aberrantes, e outros dados de não-úteis. redução de dimensionalidade pode ser útil ao analisar dados de sensores e outros Internet of Things (Internet das coisas) casos de uso. Nos sistemas da Internet das coisas, pode haver milhares de pontos de dados, simplesmente dizendo-lhe que um sensor é ligado. Armazenar e analisar que "em" dados não é útil e vai ocupar espaço de armazenamento importante. Além disso, removendo esses dados redundantes, o desempenho de um sistema de aprendizagem de máquina vai melhorar. Finalmente, a redução da dimensionalidade também ajudará os analistas visualizar os dados.

## Instância baseou

*algoritmos baseados em instância* são usados quando você quer categorizar novos pontos de dados com base em semelhanças com dados de treinamento. Este conjunto de algoritmos são por vezes referido como *alunos preguiçosos* porque não há nenhuma fase de treinamento. Em vez disso, algoritmos baseados em instâncias simplesmente coincidir com novos dados com os dados de treinamento e categorizar os novos pontos de dados com base na similaridade com os dados de treinamento.



REMEMBER

aprendizagem baseada em instância não é adequado para conjuntos de dados que têm variação aleatória, dados irrelevantes, ou de dados com valores em falta. algoritmos baseados em instância pode ser muito útil no reconhecimento de padrões. Por exemplo, a aprendizagem exemplo é utilizado na análise da estrutura química e biológica e análise espacial. Análise nas áreas biológica, farmacêutica, química e engenharia, muitas vezes usa vários algoritmos baseados instância.

## Redes neurais e aprendizagem profunda

Uma rede neural tenta imitar a forma como um cérebro humano se aproxima de problemas e usa camadas de unidades interligadas para aprender e inferir relações com base em dados observados. Uma rede neural pode ter várias camadas ligadas. Quando há mais de uma camada escondida em uma rede neural, às vezes é chamado

*aprendizagem profunda*. modelos de redes neurais são capazes de ajustar e aprender como alterações de dados. As redes neurais são usadas frequentemente quando os dados não é marcado ou não-estruturados. Um dos casos de uso chave para redes neurais é a visão de computador. (Para mais detalhes sobre redes neurais, consulte o Capítulo 1).

aprendizagem profunda está sendo alavancado hoje em uma variedade de aplicações. Auto-dirigir carros usam aprendizagem profunda para ajudar o veículo compreender o ambiente em torno do carro. Como as imagens de captura de câmeras do ambiente circundante, algoritmos de aprendizagem profundas interpretar os dados não estruturados para ajudar o sistema fazer perto decisões em tempo real. Da mesma forma, o aprendizado profundo é incorporado em aplicações que os radiologistas utilizam para ajudar a interpretar imagens médicas.

A Figura 1/3 mostra a arquitectura de uma rede neural. Cada camada dos filtros de redes neurais e transforma os dados antes de passá-lo para a próxima camada.

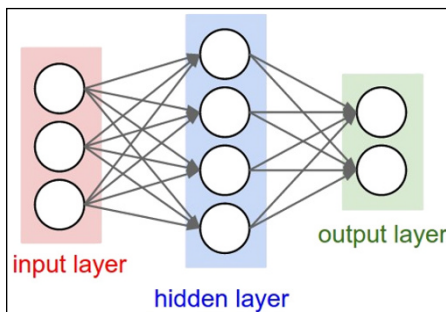


Figura 3-1: A arquitetura de uma rede neural.

## Regressão linear

*algoritmos de regressão* são comumente utilizados para análise estatística e são algoritmos de chave para uso em aprendizado de máquina. Algoritmos de regressão ajudam os analistas a relacionar o modelo entre pontos de dados.



TIP

Algoritmos de regressão podem quantificar a força de correlação entre variáveis de um conjunto de dados. Além disso, a análise de regressão pode ser útil para prever os valores futuros de dados com base em valores históricos. No entanto, é importante lembrar-se que a análise de regressão assume que a correlação relaciona-se com a causa. Sem entender o contexto em torno de dados, a análise de regressão pode levar a previsões imprecisas.

## Regularização para evitar overfitting

*regularização* é uma técnica para modificar modelos para evitar o problema de overfitting. Você pode aplicar regularização a qualquer modelo de aprendizagem de máquina. Por exemplo, você pode regularizar um modelo de árvore de decisão. Regularização simplifica modelos excessivamente complexos que são propensos a ser overfit. Se um modelo é overfit, ele vai dar previsões imprecisas quando ele é exposto a novos conjuntos de dados.



REMEMBER

Overfitting ocorre quando um modelo é criado por dados específicos definidos, mas terá capacidades de previsão pobres para um conjunto de dados generalizado.

### aprendizado de máquina baseada em regras

*algoritmos de aprendizado de máquina baseados em regras* usam regras relacionais para descrever dados. Um sistema baseado em regras pode ser contrastado a partir de sistemas de aprendizado de máquina que criam um modelo que pode ser geralmente aplicado a todos os dados de entrada. No resumo, o uso de regras são muito fáceis de entender: Se os dados X são introduzidos, fazer Y. No entanto, como os sistemas se tornam operacionalizados, uma abordagem baseada em regras de aprendizagem de máquina pode se tornar muito complexa.

Por exemplo, um sistema pode incluir 100 regras pré-definidas. Como o sistema encontra mais e mais dados e é treinado, é provável que centenas de exceções às regras possam emergir. É importante ter cuidado ao criar uma abordagem baseada em regras que ele não se torne tão complicada que ela perde a sua transparência. Pense em quão complicado seria a criação de um algoritmo baseado em regras para aplicar o código tributário.

## sistemas de aprendizagem máquina de treinamento

Através de um processo iterativo de desenvolvimento e aperfeiçoamento de um modelo, selecionando o algoritmo correto, treinamento e teste de um sistema pode começar. O treinamento é um passo crítico no processo de aprendizagem de máquina.



**TIP**

Quando você está treinando um sistema de aprendizagem de máquina, você sabe que as entradas (para renda exemplo cliente, comprando história, localização, e assim por diante), e você sabe que seu objetivo desejado (prevendo a propensão de um cliente de churn). No entanto, a grande incógnita é as funções matemáticas para transformar esses dados brutos em uma previsão rotatividade de clientes. Como o algoritmo de aprendizagem é exposto a mais e mais dados do cliente, o sistema vai se tornar mais preciso em prever a probabilidade de rotatividade de clientes.

Formação de um algoritmo de aprendizado de máquina para criar um modelo exato pode ser dividido em três etapas:

### 1. Representação.

O algoritmo cria um modelo para transformar os dados introduzidos para os resultados desejados. Como o algoritmo de aprendizagem é exposto a mais dados, ele vai começar a aprender a relação entre os dados brutos e que os pontos de dados são fortes preditores para o resultado desejado.

### 2. Avaliação.

Como o algoritmo cria vários modelos, seja um ser humano ou o algoritmo terá de avaliar e pontuar os modelos com base em qual modelo produz as previsões mais precisas. É importante lembrar que após o modelo é operacionalizada, ele vai ser exposto aos dados desconhecidos. Como resultado, certifique-se o modelo é generalizado e não overfit aos seus dados de treinamento.

### 3. Otimização.

Após o algoritmo cria e modelos contagens múltiplas, selecione o melhor algoritmo de desempenho. Como você expor o algoritmo para mais diversos conjuntos de dados de entrada, selecione o modelo mais generalizada.



**REMEMBER**

A parte mais importante do processo de formação é ter dados suficientes para que você está em uma posição para testar seu modelo. Muitas vezes, o primeiro passo na formação fornece resultados mistos. Isso significa que você quer ser necessário refinar o seu modelo ou fornecer mais dados.

Este processo não é diferente de aprender qualquer nova disciplina onde você começa com suas suposições baseadas no conhecimento incompleto. Como você aprende mais, você pode decidir se você precisa de mais dados de mais fontes. Como você ganhar mais conhecimentos a partir dos dados, suas suposições provavelmente irá mudar. Um dos valores de aprendizagem de máquina é que você não começar o processo de aprendizagem por decidir com antecedência o que a resposta para o problema será.

Quando tiver concluído o processo de treinamento, você está pronto para testar sua compreensão do domínio para ver se você tem o direito quantidade de conhecimento ou se você ainda são obrigados a recolher mais dados e aprender mais. Este é precisamente o que acontece de forma automatizada quando você desenhar um sistema de aprendizagem de máquina.

## Preparação de dados

algoritmos de aprendizado de máquina muitas vezes obter a maioria da atenção quando as pessoas discutem a aprendizagem de máquina; no entanto, o sucesso depende de bons dados.



REMEMBER

Compreender seus dados é fundamental para o seu sucesso. Se você criar um modelo baseado em dados falhos, suas previsões, obviamente, será impreciso. Além disso, você precisa pensar sobre quais os dados que devem ser incluídos em seu aplicativo de aprendizagem de máquina.

### Identificar dados relevantes

decisões de negócios precisam ser feitas com base em constante mutação dados de uma variedade de fontes. Suas fontes de dados podem incluir ambos os sistemas tradicionais de registro de dados (como o cliente, produto, transacional e dados financeiros) e os dados externos (por exemplo, mídia social, notícias, dados meteorológicos, dados de imagem ou dados geoespaciais). Além disso, muitas estruturas de dados são essenciais para a análise de informações, incluindo dados estruturados e não estruturados.

#### fontes de dados estruturados

Os dados estruturados são normalmente armazenados em bases de dados relacionais tradicionais e refere-se a dados que tem um comprimento e formato definido. A maioria das organizações têm uma grande quantidade de dados estruturados em seus centros de dados no local. Exemplos de dados estruturados incluem o seguinte:

» **dados do sensor:** Exemplos incluem rádio ID frequência (RFID)

etiquetas, medidores inteligentes, dispositivos médicos, e dados de Sistema de Posicionamento Global (GPS).

» **dados Weblog:** Quando os servidores, aplicações, redes, e assim

em operar, eles capturam todos os tipos de dados sobre a sua actividade.

» **Ponto-de-venda de dados:** Quando o caixa furto o código de barras de

qualquer produto que você compra, todos os dados associados com o produto é gerado.

» **Dados financeiros:** Muitos sistemas financeiros estão agora programa-

matic; que operam com base em regras pré-definidas que automatizam processos.

» **Dados meteorológicos:** Sensores para coletar dados meteorológicos estão sendo

implantado em cidades, regiões e para coletar dados sobre coisas como temperatura, vento, pressão barométrica e precipitação. Esses dados podem ajudar os meteorologistas criar previsões hiperlocais.

» **Clique com o fluxo de dados:** Os dados são gerados a cada vez que você clicar em um

link em um site. Estes dados podem ser analisados para determinar o comportamento do cliente e padrões de compra.

## fontes de dados não estruturados

Embora os dados não estruturados tem alguma estrutura implícita, não se segue um formato especificado. dados não estruturados ainda é amplamente subutilizados pelas empresas e oferece uma grande oportunidade para monetização. Nuvem, móvel e social têm contribuído para um aumento enorme de dados não estruturados. Exemplos de dados não estruturados incluem o seguinte:

» **Texto interno para sua empresa:** Pense em todo o texto

dentro de documentos, registros, os resultados da pesquisa e e-mails. informações da empresa realmente representa uma grande porcentagem das informações de texto no mundo de hoje.

» **Sociais de dados de mídia:** Esta informação é gerada a partir do social,

plataformas de mídia, como o YouTube, Facebook, Twitter, LinkedIn e Flickr.

» **Dados móveis:** Isso inclui mensagens de texto, notas, calendário

entradas, fotos, vídeos e dados inseridos em aplicações móveis de terceiros.

» **Imagens de satélite:** Isto inclui dados de tempo ou dos dados que

o governo captura em seu imaginário vigilância por satélite.



» **Fotografias e vídeo:** Isto inclui segurança, vigilância, e dados de tráfego.

» **Radar ou sonar de dados:** Isto inclui veículos, meteorológica, e dados oceanográficos.

## governar dados

Compreender e governar seus dados são pré-requisitos para uma utilização eficaz do aprendizado de máquina para resolver problemas reais de negócios. Haverá um nível diferente de governança quando você está treinando dados do que quando você usar esses dados em um ambiente de produção. No mundo tradicional de data warehouses ou gestão de banco de dados relacional, é provável que a sua empresa tem regras sobre como os dados precisam ser tratadas e protegidas bem compreendido. Por exemplo, no setor de varejo, é fundamental que certas disposições de segurança são colocadas em torno da informação personably identificação dos clientes. Você tem que ter certeza de que os indivíduos não autorizados não pode acessar dados privados ou restritos. Você também tem controle sobre quem tem permissão para ambos os dados visualizar e alterar os dados.



**WARNING**

Como sua organização começa a usar soluções baseadas em aprendizado de máquina para prever resultados, você deve considerar as implicações de governança de dados. Ao criar aplicativos de aprendizado de máquina, pensar sobre os três seguintes considerações de governança de dados:

» **Garantir que os dados privados não seja comprometida. No início**

de um projeto, compreender que tipos de dados serão tocados por uma aplicação de aprendizagem de máquina. Por exemplo, se o cliente processo de aplicações ou dados dos funcionários que é coberto pelas regras da indústria ou regulamentos governamentais? Se os resultados de um algoritmo de aprendizado de máquina produzir dados adicionais do cliente, esses resultados podem precisar de ser protegidos.

» **colocação de dados devem ser conduzidos por regras de governança.**

Compreender onde os dados serão fisicamente localizado e onde a aprendizagem de máquina terá lugar. Alguns países exigem que os dados cidadão ser mantido dentro do país. Outras regras e regulamentos podem proibir certos dados de mover-se para uma nuvem pública. Estes requisitos de localidade dados é importante para considerar se as aplicações mover dados para locais diferentes para executar tarefas de aprendizagem de máquina.

» **Manter a privacidade dos dados sensíveis. Entenda quem é**

permissão para ver dados que está sendo ingerida em uma aplicação de aprendizagem de máquina.

# O Ciclo de Aprendizagem Máquina

Criando uma máquina de aprendizagem aplicação ou operacionalização de um algoritmo de aprendizagem de máquina é um processo iterativo. Você não pode simplesmente treinar um modelo de uma vez e deixá-lo sozinho - alterações de dados, preferências evoluir, e os concorrentes vão surgir. Portanto, você precisa para manter seu modelo de fresco quando ele entra em produção. Enquanto você não terá que fazer o mesmo nível de formação que foi necessário quando você construiu o modelo, você não pode assumir que será auto-suficiente.



REMEMBER

O ciclo de aprendizagem de máquina é contínua, e escolher o algoritmo de aprendizado de máquina correta é apenas um dos passos. As etapas do ciclo de aprendizado de máquina são os seguintes:

- » **Identificar os dados:** Identificar as fontes de dados relevantes é o primeiro passo no ciclo. Além disso, como você desenvolver seu algoritmo de aprendizagem de máquina, pensar em expandir os dados de destino para melhorar o sistema.
- » **Preparar dados:** Certifique-se de seus dados é limpo, seguro, e governados. Se você criar um aplicativo de aprendizado de máquina com base em dados imprecisos, a aplicação irá falhar.
- » **Selecione o algoritmo de aprendizado de máquina:** Você pode ter vários algoritmos de aprendizado de máquina aplicáveis aos seus dados e desafio empresarial.
- » **Trem:** Você precisa treinar o algoritmo para criar o modelo.  
Dependendo do tipo de dados e algoritmo, o processo de formação podem ser supervisionado, sem supervisão, ou aprendizado por reforço.
- » **Avalie:** Avalie seus modelos para encontrar o melhor desempenho algoritmo.
- » **Implantar:** algoritmos de aprendizagem de máquina criar modelos que podem ser implantado para ambas as aplicações em nuvem e no local.
- » **Prever:** Após a implantação, começar a fazer previsões baseadas em dados novos, recebidas.
- » **Avaliar previsões:** Avaliar a validade das suas previsões.  
As informações que você reunir a partir de analisar a validade das previsões é então alimentado de volta para o ciclo de aprendizagem de máquina para ajudar a melhorar a precisão.



**TIP**

Após seu modelo começa a fazer previsões, começar o processo novamente, avaliando os dados que você está avaliando. É todos os dados relevantes? Há novos conjuntos de dados que podem ajudar a melhorar a precisão das previsões? Ao melhorar continuamente os modelos e avaliar novas abordagens que você será capaz de manter seus aplicativos baseados em aprendizagem de máquina relevante.

- » Compreender como a aprendizagem de máquina suporta seus objetivos
- » Centrando-se sobre o problema de negócios
- » exigindo colaboração
- » A seleção de um projeto piloto
- » Determinar o melhor modelo de aprendizagem

## Capítulo 4

# Introdução ao Aprendizado de Máquina

**voce** alcançar um nível de sofisticação das análises avançadas requer um plano e roteiro. Você não pode simplesmente contratar um grupo de cientistas de dados e esperar que eles sejam capazes de produzir resultados para o negócio.

Neste capítulo, vamos nos concentrar sobre a melhor abordagem para iniciar o processo de permitir a aprendizagem de máquina para apoiar seus objetivos de negócio. Pense em como você pode começar assim que você pode obter insights a partir dos dados gerados por sua empresa. Se você se aproxima da adoção de técnicas de aprendizado de máquina de uma forma sistemática, você estará em uma boa posição para antecipar as mudanças em seu mercado e mudanças na forma como os clientes esperam para fazer negócios com você.

## Entendendo como Machine Learning pode ajudar

Antes de escolher um projeto de destino, começar ajudando gestão de negócios entender o aprendizado de máquina é tudo. Não é uma cura para tudo. aprendizagem de máquina é uma abordagem que permite que você use

algoritmos para criar modelos baseados em dados. Portanto, é importante para definir as expectativas. No Capítulo 5, discutimos os tipos de habilidades que sua equipe precisa. Enquanto você certamente vai ter especialistas, como os cientistas de dados, é importante que os analistas de negócios e estrategistas de negócios entender como aprendizagem de máquina pode ser aplicado para o negócio para resolver alguns problemas muito complexos. A abundância e variedade de dados pode fornecer o negócio com uma arma valiosa para ajudar o seu negócio crescer e mudar.

## Concentre-se sobre o problema de negócios

Quando você começa a aplicar técnicas de aprendizado de máquina para apoiar sua estratégia de negócio, você tem que entender três fundamentos:

### » Qual é o problema de negócio que você está tentando resolver?

Certifique-se de que você tem uma boa compreensão da natureza do problema que você está tentando resolver. Você pode ver mudanças em sua receita, ou talvez nos tipos de produtos de seus clientes estão comprando. Você entende por que seus clientes estão comprando? Você entende como as mudanças no mercado estão afetando sua capacidade de satisfazer os clientes? Você tem um monte de informações sobre os clientes, mix de produtos, e seu mercado em geral, mas é necessário realizar uma análise mais profunda para que você esteja preparado para o futuro. Talvez você está considerando oferecer um novo produto para sua base de clientes tradicional. Você precisa entender como os novos produtos irão impactar sua receita durante o próximo ano.

### » Onde estão os recursos de dados ocultos que você pode tomar vantagem de compreender melhor suas oportunidades e ameaças?

Sua organização provavelmente tem muito mais informações sobre o seu negócio do que você imagina. Há registros de apoio ao cliente que pode lhe dar insights sobre as questões que confrontam seus clientes. Dados vai lhe dar insights sobre a quantidade de tempo que leva para reparar um problema. Alguns dados também é armazenada em texto que indica que os clientes estão procurando no futuro. Enquanto existe esta dados, ele pode nunca ter sido usado para fazer sentido do seu negócio. Ironicamente, você já pode ter todos os dados que você precisa para começar

avaliar o seu futuro. Todos esses dados tem o potencial para ajudá-lo a olhar para além do óbvio e antecipar o futuro.

» **Como você pode se preparar para obter seus dados em ordem?**

O desafio é ter certeza de que você tem seus dados pronto para executar o tipo de análise analítica de modo que você pode aprender com os dados que você tem. Você está usando as fontes de dados adequadas que são os mais atualizados? Você colocar os dados em um formulário que é utilizável? Você está protegendo a identidade dos dados privados dos seus clientes? Você está selecionando as melhores fontes de dados de terceiros que irão colocar seus próprios dados no contexto de sua indústria?

Enquanto aprendizado de máquina tem captado a atenção do mercado de tecnologia e negócios, você quer ter certeza de que você selecionar a abordagem e ferramentas que melhor combinar com o problema que você precisa resolver. Haverá diferentes abordagens, dependendo da sua indústria e do tipo de dados que você está lidando com eo tipo de resultados que você está olhando para conseguir.



**REMEMBER**

Para muitas organizações, sendo capaz de compreender os padrões ocultos dentro de seus dados oferece uma vantagem enorme potencial. A maioria das empresas têm dados importantes armazenados em silos em diferentes unidades de negócios. Alguns dos dados importantes podem ser encontradas em fontes de mídia social. Os dados também podem ser encontradas em fontes de dados não estruturados, como documentos relacionados com novos resultados da investigação. Os dados também é encontrada em fontes de semi-estruturado, tais como sensores e sistemas baseados em IdC.

Sua primeira tarefa é determinar que fontes de dados e tipos de dados são os mais adequados para resolver o seu problema. Depois de entender isso, você estará em uma boa posição para determinar quais algoritmos serão utilizadas para criar os modelos mais apropriados. Embora existam centenas de casos de uso para ilustrar como usar algoritmos de aprendizado de máquina para resolver problemas específicos, nós damos-lhe três exemplos desta seção.

## **Trazendo silos de dados em conjunto**

Você está em um mercado competitivo com um monte de empresas emergentes determinados a perturbar o mercado. Portanto, você tem que descobrir uma maneira de entender as mudanças sutis para as preferências e necessidades dos clientes. Enquanto você é diligente sobre a realização de pesquisas com clientes e responder às reclamações dos clientes, este

informação tende a ser silos entre unidades de negócios. Cada organização que se envolve com os clientes compreende uma visão diferente do cliente. E se você pudesse ter uma perspectiva mais ampla de todos os pontos de contato e interações com os clientes? O que poderia dizer-lhe sobre as preferências do seu cliente que você não sabia? Muitas dessas unidades de negócios vai lidar com diferentes linhas de produtos com diferentes compradores.



#### TIP

Com aprendizagem de máquina, você pode reunir uma variedade de fontes de dados internas e externas e criar um modelo que ajuda a padrões de desvendar e anomalias que impactam o que você oferece aos clientes e como você oferecer produtos e serviços.

Imagine uma cadeia de vestuário que tinha os dados e aplicado o algoritmo mais apropriado para obter uma compreensão das mudanças de expectativas dos clientes - o que eles estão felizes com e onde eles estão insatisfeitos. Esta informação fornece insights sobre os padrões de compra em mudança. É a base de clientes em crescimento? Os clientes estão começando a deixar existente? Quais são os dados demográficos de novos clientes? Os novos clientes que compram os mesmos produtos da mesma maneira como os clientes existentes? As empresas de sucesso têm a capacidade de aproveitar verdadeiramente os seus dados por quebrar silos de dados através das fronteiras organizacionais. empresas disruptivas são ágeis e compreender o valor de seus dados em crescer sua base de clientes e receita. Ganhando insights e indicadores precoces de dados pode transformar um problema em uma oportunidade.

## Evitar problemas antes que aconteça

As grandes cidades muitas vezes têm recursos limitados para lidar com a complexidade. Alguns problemas podem prejudicar a sua capacidade de reagir aos problemas que têm o potencial para superar os governos. Os problemas de trânsito pode causar engarrafamentos, levar a acidentes, causa poluição e tornar as cidades inabitável. Quando há incidentes como uma inundação ou um colapso da ponte, os serviços de apoio da cidade precisa estar preparado para agir antes que as populações são afetados drasticamente. Uma cidade inabitável tem um tempo difícil atrair novas empresas para se mudar.



#### REMEMBER

Modelando padrões de tráfego através da ingestão de dados meteorológicos, dados sobre rotas de tráfego alternativos e mídias sociais, por exemplo, pode ajudar a gestão da cidade de alerta para que eles possam alertar os cidadãos e redirecionar o tráfego longe de zonas de perigo. Ser capaz de antecipar os problemas antes que eles aconteçam pode melhorar as condições que tornam uma cidade vibrante e evitar a perda de vidas e bens. Como você faz isso?

aprendizado de máquina pode aprender os padrões e condições que podem mudar os padrões de tráfego em um ritmo que a mente humana não pode.

## Primeiros focada no cliente

Inovações acontecem muitas vezes quando uma empresa começa a entender que existe uma melhor maneira de criar oportunidades de negócios. A única maneira de estar preparado para a mudança é ter os dados e análises que ajudam a determinar a próxima melhor ação para obter os resultados que esperamos. Buscando a resposta a um problema só funciona quando você tem uma idéia do que a resposta pode ser. Com aprendizagem de máquina, é possível encontrar soluções quando você não pode antecipar o que as respostas ou resultados será.



**TIP**

Com base no entendimento mudanças de expectativas, é possível para ajudar os clientes sabem o que querem antes que eles possam articular a necessidade. Compreender as mudanças sutis nos padrões de compra dos clientes pode ajudar a dinamizar o negócio de mudar constantemente embalagem e ofertas. Ironicamente, as empresas muitas vezes pode pegar esses dados a partir de uma variedade de fontes de dados públicos. Combinando esses dados com informações sobre os seus próprios clientes pode levar a algumas abordagens potencialmente vencedoras.

## Machine Learning requer a colaboração

Grande parte do foco na aprendizagem de máquina é a viabilidade dos modelos que são criados por cientistas de dados. Mas, para os modelos para ser capaz de prever os resultados do negócio, tem que haver uma colaboração rigorosa com o negócio. Linha de Negócios líderes (LOB) são mais capazes de compreender a dados importantes que é utilizado para analisar o negócio. No entanto, muitas vezes eles têm um viés sobre o que é mais importante para os clientes e os dados que é mais importante. É fundamental que as equipes de ciência os dados e os dados do Analytics descobrir novas fontes de dados que podem melhorar a capacidade das empresas para descobrir os padrões e tendências ocultas. O nível apropriado de colaboração entre unidades de negócio, liderança corporativa, e os cientistas de dados pode criar valor que conduz à verdadeira diferenciação e mudança significativa.



## A execução de um Projeto Piloto

Depois de ter uma ideia de quais os tipos de problemas que você pode começar a resolver com a aprendizagem de máquina, você pode estar pronto para começar a experimentar. Não espere que você pode fazer tudo com a contratação de alguns cientistas de dados e tê-los experiência em isolamento. Você precisa criar uma colaboração entre esses analistas de negócios, executivos, estrategistas, cientistas de dados e analistas.

Nesta seção, nós damos-lhe os passos para ajudá-lo a executar com sucesso um projeto piloto.

### Passo 1: Definir uma oportunidade para o crescimento

Comece com um problema que pode ser vinculado a um resultado de negócios. Mas não se chegar à frente de si mesmo. Certifique-se de que você selecione um pequeno problema onde você pode facilmente identificar os dados que você já tem e as fontes de dados que você pode obter.



REMEMBER

O que é a oportunidade que a sua empresa pode tirar proveito de fazer crescer o negócio? Talvez você tenha identificado um produto tradicionalmente popular que não está vendendo bem. Por exemplo, qual é a embalagem ideal de seus produtos que irão aumentar as vendas no futuro? Ao compreender e modelar os dados, você pode entender como os dados que você tem ajuda a prever o melhor conjunto de opções que atenda às necessidades de seus clientes. Seu piloto selecionado também é uma ferramenta de marketing para demonstrar ao negócio que você pode antecipar as futuras necessidades dos clientes.

### Passo 2: Realização de um projeto piloto

Começar a realizar o seu piloto com a sua ideia concreta a partir do Passo 1. Certifique-se de explicar o propósito do projeto e do tipo de dados que você está usando. Bons projectos-piloto são um subconjunto de um problema maior que você está tentando resolver para o negócio. Se o piloto for bem sucedida, você já ter colocado seus objetivos no contexto. Você terá ganho conhecimentos para definir seus próximos passos.



REMEMBER

Você vai aprender uma quantidade enorme de este projecto-piloto. Enquanto você certamente vai aprender muito com um piloto bem sucedido, você pode aprender ainda mais de um piloto falhou. O que você pode aprender sobre os padrões de compra dos clientes? você pode determinar as maneiras que os clientes

comprar seus produtos hoje e como isso começou a mudar? Estes novos padrões que emergem a partir dos dados e ser capaz de prever o que isso pode significar para o seu negócio pode ajudar a sua estratégia de transformação.

## Passo 3: Avaliação

Vamos supor que você aprende algumas coisas interessantes de conduzir seu piloto (ver Passo 2). Você está vendo alguns padrões interessantes emergem sobre os clientes e suas necessidades futuras. Como os resultados diferem da maneira que você realizar negócios hoje? Sua gestão tem feito certas suposições sobre seus clientes eo que eles querem. Será que os resultados de seu piloto indicam que você estava fazendo suposições que não estavam em linha com os resultados do piloto?



**TIP**

Você pode se surpreender que quando você remove preconceitos, seus resultados diferem significativamente do que você pensou que iria encontrar. Este é um dos grandes benefícios da aplicação de aprendizagem de máquina para um problema de negócios - você será capaz de compreender o seu negócio em uma maneira muito diferente. Você pode achar que o piloto indica que as suas expectativas de clientes são muito diferentes do que os seus pressupostos e preconceitos. À medida que adiciona novas fontes de dados, as mudanças nos requisitos dos clientes tornam-se mais definido. Estas respostas, então, alimentar em seu planejamento de negócios e pode permitir que a sua empresa a se mover mais rapidamente para tentar novas abordagens que podem impactar positivamente a receita.

## Passo 4: ações Próxima

O benefício do projeto-piloto é que ele começa a dar-lhe uma compreensão de como você pode usar aprendizado de máquina e análise preditiva para entender melhor o seu negócio. Se você planejou seu piloto como o primeiro passo de uma série de projetos, você estará bem em sua maneira de aprender a partir de seus dados. Nesta fase, você quer expandir para incorporar mais dados e trazer mais líderes de negócios no processo. Selecione mais fontes de dados de diversas áreas do negócio que ajudam o seu processo de análise. Com aprendizagem de máquina, mais dados que você pode aplicar a um projeto a sua melhor chance de ganhar insights que você pode aplicar a sua estratégia de negócios.

## Determinando o melhor modelo de aprendizagem

Uma das tarefas mais complexas para a aplicação de aprendizagem de máquina para um problema de negócios é selecionar o modelo mais adequado. Selecionando o modelo mais adequado é o melhor ponto de partida para a viagem para máquina de fazer aprender uma ferramenta indispensável para prever resultados de negócios. Uma das questões mais complexas com a seleção de um modelo é certificar-se que o modelo terá um bom desempenho no futuro, quando novos dados são introduzidos.



**WARNING**

O algoritmo selecionado tem que ser generalizado o suficiente para que ele possa ser preciso com novos dados. Se o algoritmo for demasiado firmemente amarrado a um conjunto existente de dados, este tipo de overfitting vai causar problemas no futuro. Portanto, quando você seleciona um algoritmo, começar por certificando-se de que o conjunto de dados a ser utilizado é uma amostra representativa de suas informações. Seu piloto será muito mais bem sucedido se o conjunto de dados é uma amostra representativa do aspecto do negócio que você está focado em. Por exemplo, você pode começar a selecionar um algoritmo, selecionando um conjunto de dados de amostra que é bem conhecido na sua organização. Como próximo passo, você pode adicionar um conjunto de uma fonte totalmente diferente que poderiam ser relevantes para a sua hipótese de dados. Como é que o algoritmo que você selecionou prever os resultados de ambos os dados bem compreendidos estabelecidos e os novos dados set?

### Ferramentas para determinar a seleção algoritmo

Definitivamente, não é fácil de selecionar o algoritmo que é mais adequado para os seus dados e o seu desafio. Felizmente, o mercado está começando a reconhecer que, a fim de avançar, ferramentas precisam existir para ajudar com a seleção algoritmo. Como você escolher o modelo certo? É um problema difícil.



**WARNING**

Enquanto overfitting pode ser um problema, um problema mais sério é que os modelos perder precisão ao longo do tempo. Portanto, você tem que treinar continuamente o modelo como as alterações de dados. Selecionando o algoritmo de direito pode ser melhor realizado ao automatizar a seleção de um algoritmo. Tomemos o exemplo de um algoritmo de classificação. Há cerca de 40 diferentes algoritmos classificadores. Estes algoritmos diferentes podem ser combinados, dependendo da abordagem do cientista de dados está a utilizar. Portanto, você pode ter centenas de combinações para você escolher. Se os cientistas seus dados necessário para testar algoritmos potencialmente válidos, que poderia levar um longo tempo para escolher

os melhores. Usando uma ferramenta de automação permite que os cientistas a determinar mais rapidamente a melhor combinação de algoritmos que irá fornecer a maior pontuação e o melhor ajuste para seus dados.

ferramentas de automação são importantes não apenas por causa da complexidade dos algoritmos mas também porque você tem que ter certeza que os algoritmos que você selecionar para construir seus modelos não terá impacto sobre a latência de dados e consistência dos dados.

## **seleção da ferramenta que se aproxima**

Uma variedade de ferramentas de código aberto têm a intenção de ajudar os cientistas de dados, selecione o algoritmo certo. Essas ferramentas são muitas vezes ligada diretamente à linguagem (Python, R, Java, e assim por diante) a ser utilizado. Por que os cientistas de dados usar ferramentas para a seleção algoritmo? Muitos modelos de aprendizagem de máquina diferentes podem todos ser útil na resolução de problemas. Se um cientista de dados pode experimentar com diferentes algoritmos, ele será capaz de melhorar a capacidade dos modelos para prever resultados e criar modelos que irá escalar.



- » Identificar as competências para a sua equipa
- » Encontrar recursos para ajudar a aprender mais sobre a aprendizagem de máquina

# Capítulo 5

## Aprendizado de máquina Skills

**E**u tenho um bom senso das complexidades e os benefícios de alavancar a aprendizagem de máquina para resolver problemas de negócios. Você sabe sobre isso, mas não sabe como armar sua equipe com as habilidades certas, incluindo linguagens e ferramentas. Neste capítulo, nós lhe fornecer uma compreensão das tecnologias que ajudam sua organização alavancar com sucesso os benefícios da aprendizagem de máquina para apoiar os objetivos de negócio da sua organização.

### Definindo as habilidades que você precisa

Sua equipe precisa de uma variedade de ferramentas para aplicar com sucesso a aprendizagem de máquina para resolver alguns dos seus problemas de negócios mais complexos. À primeira vista, você pode esperar que você pode empregar uma grande equipe de cientistas de dados. No entanto, a realidade é que é difícil encontrar os cientistas de dados que você precisa para mover sua empresa para a frente rapidamente. Há simplesmente não é suficiente cientistas de dados qualificados. E, porque este talento é difícil de encontrar, você terá que pagar altos salários para os cientistas de dados que você descobrir. A resposta é que você precisa pensar de forma diferente sobre como você pessoal um departamento focado em inovando com a aprendizagem de máquina. Você pode se concentrar os cientistas de dados na construção de modelos que podem ser usados

por analistas de dados experientes. Ao mesmo tempo, você pode começar a selecionar as ferramentas de última geração que podem ajudar um analista inteligente ser treinados para alcançar muitas das técnicas de aprendizado de máquina. Existem muitos cursos de treinamento online que podem ajudar a educar sua equipe.

Nesta seção, nós damos-lhe dez áreas de habilidade que recomendamos ser seu foco. Cada uma dessas áreas tem muitos elementos. Portanto, certifique-se de que sua equipe mergulha profundamente nas áreas que impactam a capacidade da organização para suportar o negócio.

## Entenda o que ferramentas estão disponíveis

Quais são as características de liderança para apoiar objetivos da sua empresa com a **aprendizagem de máquina? Não há 1 única ferramenta ou técnica que você pode usar para** aprendizagem de máquina; você pode usar uma variedade de ferramentas. Você deve gastar algum tempo experimentando com diferentes abordagens que melhor combinam com o problema que você está tentando resolver. Existem melhores práticas que podem ajudar neste processo de seleção da ferramenta.

## Aprender línguas

Uma série de linguagens populares pode ser útil em avançar com a aprendizagem de máquina. A popularidade de linguagens muda ao longo do tempo por isso é muitas vezes útil para aprender mais de um idioma. Linguagens como Python, R, Java e C ++ são fundamentais para avançar com a aprendizagem de máquina. Ferramentas como o Linux, Hadoop, Spark, e serviços de nuvem são necessários para operar em um ambiente onde você está investindo em aprendizado de máquina.

## Explorar algoritmos

Você precisa entender os inúmeros algoritmos que serão úteis no aprendizado de máquina. Um bom cientista de dados terá profunda compreensão dos métodos de probabilidade e estatística, porque estes são frequentemente utilizados na criação de modelos de aprendizado de máquina eficazes. algoritmos de chave que entram em jogar para a aprendizagem de máquina incluem a criação do modelo para determinar os padrões e correlações e clusters a partir dos dados. Para mais detalhes sobre algoritmos de aprendizado de máquina, consulte o capítulo 3.

## **Escolha modelos apropriados**

É importante que você aplique os algoritmos de aprendizado de máquina certos para resolver o problema na mão. Um aumento do número de embalagens de algoritmos de aprendizagem automática existir por meio de APIs, incluindo faísca MLlib, H2O, e TensorFlow. Uma das habilidades mais importantes para os desenvolvedores é entender qual algoritmo é o melhor ajuste para o problema. Por exemplo, um modelo de regressão linear se encaixa o problema quando você está tentando entender como dois pontos estão relacionados. Por outro lado, se você está lidando com a compreensão do conteúdo de imagens, você pode querer explorar TensorFlow. Muitas técnicas de aprendizado de máquina combinar uma variedade de problemas de aprendizagem. O cientista de dados precisa ser capaz de determinar qual algoritmo e bibliotecas fazem mais sentido.

## **Compreender o valor de probabilidade e estatística**

Um grande número de algoritmos de aprendizagem são baseados em probabilidade e estatística. Naive Bayes, modelos de mistura de Gauss, e Hidden Markov Models são alguns dos métodos que são importantes para entender.

## **Entenda gerenciamento de dados**

Cientistas de dados também tem que entender os dados que está sendo usado. Qual é a fonte dos dados? É que fonte confiável e rastreável? Será que as fontes que você traz em conjunto para resolver um sentido problema make? Neste caso, o programador ou dados cientista precisa de trabalhar em colaboração com a empresa para vet as fontes de dados.

## **Avaliar a limpeza de suas fontes de dados**

Como é bom suas fontes de dados são fará a diferença entre o sucesso eo fracasso de seus projetos de aprendizagem de máquina. Você precisa entender as origens de seus dados e certifique-se de que eles são respeitáveis. Você também precisa determinar se você está selecionando uma combinação de fontes de dados que fazem sentido quando postos juntos.



## A LISTA DE VERIFICAÇÃO: construir sua equipe

Como você deve planejar a sua equipe de ciência de dados? Não importa o tamanho de sua empresa é, há algumas características comuns que você vai fazer sucesso. Lembre-se, você está construindo uma equipe para resolver um problema de negócios. A equipe pode ser uma ou duas pessoas que fazem de tudo, ou em uma grande empresa, você pode ter uma pessoa para cada conjunto de habilidades. Provavelmente, você não **vai encontrar uma pessoa com todas essas habilidades (o que nos referimos como um *unicórnio*), mas aqui está** uma lista de verificação que ajuda você a começar:

- Construir uma equipe com uma mistura de habilidades. Você quer ter certeza de que você está equilibrando os membros da equipe técnica com os membros de negócios.
- Escolha um cientista de dados principal, que é bem versado em ambos programação e princípios arquitetônicos. Além disso, o indivíduo deve ter comprovada capacidade de liderança, a fim de direcionar a equipe para executar em objetivos de negócios.
- Traga um analista de negócios que conhece a sua indústria, bem como a sua empresa.
- Certifique-se de um membro da equipe pode contar uma história a partir dos dados. Esta habilidade é diferente de interpretar os dados ou a compreensão dos dados; ele está usando os dados para enquadrar uma discussão ou provocar uma ação.
- Selecione líderes empresariais representativas que entendem o que eles precisam para ganhar com o projeto.
- Adicionar especialistas no assunto para a equipe que realmente compreender os detalhes de como os processos de trabalho ea natureza dos dados. Esses especialistas devem colaborar com um engenheiro de dados que sabe como capturar e processar os dados.
- Encontrar consultores quando necessário, que podem ajudar a treinar a equipe em novas linguagens ou novas ferramentas que suportam os objetivos do projeto.
- Trazer especialistas para áreas técnicas específicas, onde você não tem talento in-house.

Se você trabalha em uma grande empresa, você pode ter uma variedade de pessoas dentro de sua organização que são adequadas para as tarefas. Neste caso, você quer ter certeza de que você tem bons líderes que podem criar um ambiente colaborativo. Se você estiver operando em uma pequena empresa, selecione os membros da equipe que realmente entendem os fundamentos da sua organização e seus objetivos. Use os membros da equipe técnica que você já tem e complementar com especialistas da indústria, que será o mentor sua equipe.

## Entender como parte do trabalho em conjunto

A linha inferior é que, com a aprendizagem de máquina você está construindo uma aplicação baseada em um resultado de negócios. Portanto, você precisa entender como todos os elementos do software e infra-estrutura apoiar esses resultados. Como os elementos se encaixam e se comunicar uns com os outros para formar um sistema? Como você cria um ambiente que se adapta à medida que mais dados e mais lógica são adicionados? Você precisa entender que você está construindo um sistema que requer testes, gestão, documentação, e assim por diante.

## Compreender o ciclo de vida de dados

Um dos grandes benefícios da aprendizagem de máquina é o fato de que ele requer uma ingestão constante de novos dados, a fim de ser capaz de fazer previsões precisas. Portanto, você precisa entender que a aprendizagem de máquina não é uma tarefa única. Em vez disso, a aprendizagem de máquina é um continuum. Quanto mais precisa e abundante seus dados, melhor seus resultados.

## Identificar novos casos de uso

aprendizado de máquina pode ser útil em muitas indústrias diferentes e muitas funções diferentes. Explorando máquina de aprender com os pilotos a produção irá ajudá-lo a obter insights sobre novos usos. Pode haver muitas outras áreas dentro de sua empresa, que podem se beneficiar do tipo de análise preditiva que a aprendizagem de máquina pode proporcionar.

## obtendo Educated

Porque a aprendizagem de máquina é um mercado emergente, há uma grande demanda por pessoal qualificado para ajudar os esforços de organizações de apoio. Está se tornando claro que as empresas não podem esperar para encontrar todos os profissionais qualificados de que precisam. Isto significa que há uma grande oportunidade para profissionais de TI para o seu jogo e se tornar especialistas em ciência de dados e técnicas de aprendizado de máquina. Felizmente, há uma série de recursos lá fora, que pode ajudá-lo a aprender. Nesta seção, nós fornecemos uma lista de recursos que estão disponíveis para lhe dar um grande começo.

## Medium: Inside Machine Learning

Este site dá-lhe artigos profundo mergulho em uma ampla gama de tópicos de aprendizado de máquina. Desde previsões meteorológicas para robôs, você pode explorar os principais estudos de caso de aprendizado de máquina e obter insights de especialistas da indústria. Visita [medium.com/inside-machine-learning](https://medium.com/inside-machine-learning). Para maiores informações.

## CognitiveClass.ai

Visita <https://cognitiveclass.ai> para construir a ciência de dados e habilidades de computação cognitiva para livre hoje. As aulas são baseadas em uma iniciativa da comunidade IBM. Os cursos incluem "Máquina Aprender com Apache SystemML."

## aprendizagem on-line Coursera

Coursera é uma plataforma de aprendizagem on-line que oferece cursos e graus em uma variedade de áreas, incluindo a aprendizagem de máquina. Ele trabalha com universidades para oferecer mais de 2.000 cursos. Inscreva-se hoje em [www.coursera.org/learn/machine-learning](https://www.coursera.org/learn/machine-learning).

## cursos Udacity no aprendizado de máquina

Udacity é uma organização com fins lucrativos de ensino que oferece mooc on-line (MOOCs). Você pode encontrá-lo em [www.udacity.com/course/intro-to-machine-learning--ud120](https://www.udacity.com/course/intro-to-machine-learning--ud120).

## Galvanizar

Immersive currículo de ciência de dados inclui um mergulho no aprendizado de máquina e trabalhando em problemas reais na classificação, regressão e de agrupamento, utilizando conjuntos de dados estruturados e não estruturados. Os alunos descobrem bibliotecas como scikit-learn, NumPy, e SciPy, e utilizar estudos de caso do mundo real para erradicar compreensão dessas bibliotecas para aplicações do mundo real. Saiba mais em

[www.galvanize.com/san-francisco/data-science](https://www.galvanize.com/san-francisco/data-science).

## cursos EDX

EDX é um provedor de MOOC. Abriga cursos de nível universitário on-line. Alguns dos cursos são ainda oferecidos sem nenhum custo. Visita [www.edx.org / curso /-ciência máquina-learning-data-analyticscolumbia-ds102x-1](https://www.edx.org/courses/columbia/columbia-ds102x-1) para saber mais sobre a "Máquina Aprender para a Ciência dados e análises" curso on-line.

## MIT OpenCourseWare

MIT criou um site que inclui todos os seus cursos. Ele é oferecido sem nenhum custo para os participantes. Você pode aprender mais sobre a aprendizagem de máquina em <http://bit.ly/1tP7pPU>.

## Blog Google Research

pesquisadores do Google publicar uma variedade de artigos sobre temas relacionados com a aprendizagem de máquina e aprendizagem profunda. Você pode aprender mais sobre o aprendizado profundo aqui: [research.googleblog.com/2016/01/teachyourself-deep-learning-with.html](http://research.googleblog.com/2016/01/teachyourself-deep-learning-with.html).

## Kaggle Wiki

O Kaggle Wiki público é um recurso para as estatísticas de aprendizagem, aprendizagem de máquina, e outros conceitos de ciência de dados. Ele oferece tutoriais, bem como uma plataforma para competições de ciência de dados. Visita [www.kaggle.com/wiki/Home](http://www.kaggle.com/wiki/Home) hoje.

## KDNuggets

KDNuggets é um local popular que fornece uma grande quantidade de informações sobre análise e uma variedade de informações sobre a ciência de dados. Confira o conteúdo em [www.kdnuggets.com](http://www.kdnuggets.com)

## Ciência Central de Dados

Dados Ciência Central é um site online para grandes praticantes de dados. Ele inclui uma plataforma de comunidade com fóruns técnicos para a troca de informações e suporte técnico. Dirija-se a [www.datasciencecentral.com](http://www.datasciencecentral.com) Para maiores informações.

## Recursos IBM-recomendados

A comunidade de aprendizagem de máquina IBM pode fornecer fontes para adicionar ao seu conhecimento aprendizagem de máquina. Para mais informações, visite estes sites:

- » **ibm.com/machinelearning:** Veja como as empresas estão usando aprendizagem de máquina para enfrentar os desafios e buscar novas oportunidades.
- » **ibm-ml-hub.com:** Obter know-how prático de forma rápida e poderosamente aplica a aprendizagem de máquina para começar a transformar o seu negócio.
- » **ibm.com/datascience:** Pesquisar os recursos que melhor atender às suas necessidades e aprender a colaboração está permitindo que as equipes de ciência de dados de inovar com o tempo rápido de valor.
- » **datascienceforall.com:** Se você é um codificador interessado nas últimas capacidades open-source ou um analista à procura de ferramentas de arrastar-e-soltar para colaborar em projetos de ciência de dados e mover-se rapidamente, visitar a comunidade científica de dados para encontrar as últimas melhores práticas e recursos para ajudá-lo a ter sucesso.
- » **datasciencemeetups.com:** Mantenha-se atualizado sobre as últimas meetups em sua área, ou participar de um encontro virtual com dados especialistas em ciência e partilha.

Você também pode usar a mídia social para ficar conectado com o mundo da ciência de dados. Visite estas duas comunidades:

- » **Facebook:** [www.facebook.com/IBMDDataScience](https://www.facebook.com/IBMDDataScience)
- » **Twitter:** [twitter.com/IBMDDataScience](https://twitter.com/IBMDDataScience) ou @ IBMDDataScience

# Usando Machine Learning para oferecer soluções para problemas de negócios

M computação de mídia social para aplicações financeiras complexas, aprendizado de máquina, inteligência artificial e controle de qualidade, análise de sentimentos, controle e prever resultados de dados complexos, e até mesmo transformar a forma como diferentes empresas podem operar. Ser capaz de correlacionar os dados para detectar padrões e anomalias podem ajudar uma organização a prever resultados e melhorar as operações. Há inúmeros exemplos em quase todos os setores. Neste capítulo, nós damos-lhe alguns exemplos de como a aprendizagem de máquina pode ser aplicada para resolver problemas complexos de negócios.

## Aplicando Machine Learning à saúde do paciente

Um dos maiores problemas no tratamento de pacientes é que as drogas muitas vezes afetam indivíduos de forma diferente. Alguns medicamentos podem causar efeitos colaterais terríveis para um paciente ao mesmo tempo que uma eficaz

tratamento para um paciente diferente. Um paciente pode ter condições médicas adicionais que podem causar uma reação a um tratamento. Idade e sexo também podem afetar a eficácia de uma droga. Demasiadas vezes os médicos têm de recorrer a tentativa e erro para encontrar o tratamento certo.



**TIP**

Uma solução para selecionar o tratamento mais eficaz é construir um modelo de aprendizagem de máquina com base na classificação e regressão algoritmos. O modelo de classificação é necessário prever o impacto do medicamento com base em resultados de testes conhecidos e as condições do paciente. O modelo de regressão é então usado para prever as mudanças na condição do paciente quando ela toma um determinado medicamento. A criação deste modelo usando dados ajuda a fornecer pesquisadores com uma compreensão de como uma população de pacientes reage historicamente a várias drogas. Como o modelo é construído e treinado, será capaz de determinar a probabilidade de que uma determinada droga será mais eficaz para um paciente.

Se o modelo está online, ele continuará a evoluir à medida que mais dados de pacientes são adicionadas. Uma solução pode ser construída para incluir uma interface de conversação utilizando interfaces cognitivas Programação de Aplicações (API). Desta forma, um médico pode interagir com o modelo e pedir uma série de perguntas para garantir que o tratamento certo é fornecido com menos efeitos colaterais.

## Aproveitando a Internet das coisas para criar resultados mais previsíveis



**TIP**

modelos de aprendizado de máquina são uma aplicação ideal para a Internet das Coisas (Internet das coisas). A primeira coisa a entender sobre análises sobre os dados da Internet das coisas é que envolve conjuntos de dados gerados por sensores. Estes sensores são agora barato e sofisticado o suficiente para suportar uma variedade aparentemente infinita de aplicações. Os dados gerados pelos sensores contém uma estrutura específica e é, portanto, ideal para a aplicação de técnicas de aprendizagem automática. Embora os dados em si não é complexa, muitas vezes há uma enorme quantidade de dados produzidos. Ao usar esses dados de sensores, juntamente com falhas conhecidas, algoritmos de aprendizado de máquina pode construir modelos para prever futuros problemas mecânicos. O modelo que incluem dados sobre os indicadores ótimas de uma linha de base de uma máquina bem executar, bem como pontos de dados a precederam um fracasso. Como o modelo é treinado,

## COMO costumava ser feito

Machinery precisa ser gerenciado, mantidos e monitorados regularmente para assegurar o controle de qualidade e desempenho eficaz. Tomando equipamentos offline para manutenção desnecessários significa o tempo de inatividade. Da mesma forma, equipamento funcionando até falhar irá resultar em interrupções não programadas e resultados potencialmente catastróficos. Portanto, as organizações quer a capacidade de detectar potenciais problemas e corrigi-los antes que eles possam causar tempo de inatividade.

Alcançar esse nível de manutenção preventiva não tem sido fácil. Com métodos de diagnóstico tradicionais, você pode entender o que aconteceu no mês passado ou mesmo no dia passado. As empresas de manufatura foram os primeiros a adotar a tecnologia de sensores para monitorar como o equipamento bem estava operando. As empresas típicas maneira seria monitorar a saída de sensores foi determinar se eles foram combinando a saída antecipada. No entanto, a fim de evitar falhas, é importante para antecipar e prever falhas antes que possam causar danos.

Enquanto equipamento foi equipado com sensores ao longo de décadas, não havia nenhuma maneira fácil de agregar os dados criados por sensores. Com os avanços na rede e o advento de computação barato nuvem e armazenamento, é agora possível agregar esses dados do sensor. Com o advento de técnicas analíticas avançadas, é possível capturar as informações geradas por sensores e aplicar técnicas de aprendizado de máquina para prever quando uma máquina é provável que falhe.

## Proativamente Respondendo a questões de TI

operações de TI têm sido sempre complicado por causa da variedade de diferentes dispositivos de rede, servidores, aplicativos, sistemas de armazenamento, terminais, e assim por diante. Cada sistema tem suas maneiras originais de gestão dos seus componentes. À medida que novas versões do software são implementadas, atualizações de configuração pode ser necessária para manter o sistema funcionando como esperado. Este é o caminho normal que os sistemas precisam interagir de forma a manter um estado estável. Muitas vezes, um único erro em uma área pode levar a uma queda enorme, que pode ser difícil determinar a causa original de um problema - apesar do fato de que há instrumentação significativa dentro do centro de dados.



Uma organização típica pode implantar uma dúzia de diferentes ferramentas de monitoramento para tentar acompanhar a saúde de seus sistemas. Essas ferramentas de monitoramento capturam uma enorme quantidade de dados sobre os sistemas que estão monitorando. No entanto, um desafio fundamental é interpretar o grande volume de dados do sistema e o fato de que os dados estão contidos em toras. Para entender os dados, os registros devem ser compreendidos. Além desse registro e sistema de dados, dados valiosos também podem ser encontrados em registros de problemas que incluem texto que descreve um problema ou dados a partir de sistemas de gestão de desempenho de aplicações.



**TIP**

Aplicação de algoritmos de aprendizado de máquina a esses dados operações de TI complexa permite que as organizações respondam de forma proativa a potenciais problemas de TI. Tradicionalmente, correlação de eventos tem sido usado para procurar por padrões em dados de desempenho. Há momentos, no entanto, quando a correlação sozinho pode ser enganosa. Portanto, para obter melhor precisão, os cientistas de dados estão começando a fragmentação algoritmos de aprendizado de máquina para identificar anomalias de eventos. O valor da aplicação de aprendizagem de máquina é que ele pode criar um modelo baseado em um conjunto complexo de dados criados dentro do centro de dados, incluindo alertas, logs e instrumentação ou sensores. O algoritmo de aprendizado de máquina cria um modelo baseado em todos os dados relevantes. O modelo pode compreender as dependências entre os vários elementos que compõem o ambiente. O modelo também pode ajudar a identificar padrões de métricas de desempenho ideal e que ao comparar o estado atual do meio ambiente. À medida que mais dados são adicionados, o modelo pode ser continuamente atualizado.

## Proteger contra fraudes



**WARNING**

Deteção de fraudes é um jogo de gato e rato. Maus atores estão se tornando cada vez mais sofisticados em perpetrar fraude. À medida que mais e mais clientes usam serviços online, o potencial de fraude aumentou dramaticamente. Além disso, processadores de pagamento querem ter certeza de que os clientes tenham uma transação livre de atrito e não quer bloquear pagamentos legítimos. Muitas empresas estão descobrindo que a única abordagem que pode ajudar a parar a fraude é a utilização de software, com base em algoritmos de aprendizagem de máquina. Um modelo treinado pode identificar uma anomalia antes de um evento de fraude seja perpetrada. Em essência, o modelo pode identificar uma ação que está associada com uma intrusão ou uma ação não autorizada e bloquear o intruso antes que os danos possam ocorrer.



#### REMEMBER

Luta contra a fraude tornou-se um desafio complexo e leva a combinação de uma variedade de técnicas. técnicas lineares, redes neurais e aprendizagem profunda são utilizados em conjunto, a fim de detectar o comportamento fraudulento (para mais detalhes, veja os capítulos 1 e 3). algoritmos lineares têm sido usados por um longo tempo para separar as actividades válidas a partir de outras práticas fraudulentas. No entanto, um algoritmo simples não pode antecipar que o criminoso vai mudar constantemente suas técnicas. É difícil ficar um passo à frente da atividade criminosa.

Porque algoritmos lineares por conta própria não pode manchar técnicas fraudulentas avançados, mais avançados algoritmos de aprendizado de máquina são usados. Por exemplo, redes neurais e aprendizagem profunda estão sendo usados por processadores de pagamento. Os modelos de aprendizagem profundas levar em conta milhares de pontos de dados, a fim de compreender o contexto em torno de uma transação.



#### TIP

Uma organização não vai usar redes neurais ou aprendizagem profunda em isolamento. Em vez disso, **ele vai usar todas as três técnicas em conjunto a fim de realizar *modelagem de conjunto*, que** tem as suas vantagens. Por exemplo, enquanto o algoritmo linear pode perder alguma atividade fraudulenta, pode ser muito bom em pegar os esquemas mais comuns e simples. O modelo final terá de votos de cada modelo de aprendizado de máquina e aprovar ou bloquear uma transação. Este tipo de avaliação é muito semelhante a um paciente médica recebendo opiniões múltiplas dos médicos. No final, o objetivo é que as opiniões múltiplas irá produzir resultados mais precisos.



- » Incorporação de aprendizagem de máquina em aplicações
- » Tornar os dados treinados como um serviço a requisito prévio
- » Investir na aprendizagem de máquina como um serviço
- » Racionalização do aprendizado de máquina oleoduto
- » Automatizando a seleção algoritmo
- » Exigindo transparência e confiança
- » Máquina de fazer aprender um fim-de-final processo

## Capítulo 7

# Dez previsões sobre o Futuro da Aprendizagem Automática

**M** desenvolvimento tanto na indústria de software. Embora esta tecnologia avançada tem sido a aprendizagem de máquina está emergindo como um dos mais importantes. Estamos entrando em uma era em que as técnicas de aprendizado de máquina são ferramentas essenciais para a criação de valor para as empresas que querem entender o valor oculto dos seus dados. O que o futuro reserva para a aprendizagem de máquina? Neste capítulo, você explora os nossos dez principais previsões.

## Aprendizado de Máquina será incorporado na maioria das aplicações

Hoje, as técnicas de aprendizado de máquina estão começando a se tornar popular em uma variedade de ambientes especializados. As empresas estão olhando para as técnicas de aprendizado de máquina para ajudá-los a antecipar o futuro e criar diferenciação competitiva.

Nos próximos anos, você vai começar a ver os modelos de aprendizado de máquina embutidos em quase todas as aplicações e em uma variedade de dispositivos, incluindo dispositivos móveis e hubs da Internet das coisas. Em muitos casos, os usuários não sabem que eles estão interagindo com modelos de aprendizagem de máquina. Dois exemplos onde a aprendizagem de máquina modelos já estão incorporados em aplicações diárias são sites de varejo e anúncios online. Em ambos os casos, os modelos de aprendizado de máquina são muitas vezes utilizados para proporcionar uma experiência mais personalizada para os usuários.

O impacto da aprendizagem de máquina em uma variedade de indústrias será dramático e perturbador. Portanto, a aprendizagem de máquina vai mudar significativamente a forma como você faz as coisas. Por exemplo, os hospitais podem usar modelos de aprendizagem de máquina para antecipar a taxa de admissão com base nas condições dentro de suas comunidades. Admissões podem estar relacionados às condições meteorológicas, o surto de uma doença transmissível, e outras situações, como grandes eventos que ocorrem na cidade.

Estamos apenas começando a ver mais e mais modelos de aprendizagem automática embutidos em pacotes de soluções, tais como soluções de gestão de clientes e sistemas de gestão de fábrica. Com a adição de modelos de aprendizagem de máquina, esses mesmos sistemas se tornam mais inteligentes e são capazes de fornecer capacidade de previsão para aumentar o valor para a organização.

## Treinado dados como um serviço vai se tornar um pré-requisito

Um dos principais obstáculos para o desenvolvimento de modelos de aprendizagem cognitiva e máquina está treinando os dados. Tradicionalmente, os cientistas de dados tiveram que assumir os trabalhos de recolha, rotulagem, e treinar os dados. Outra abordagem é usar publicamente disponíveis conjuntos de dados ou ferramentas de crowdsourcing para coletar e dados da etiqueta. Enquanto ambas as abordagens funcionam, eles são demorado e complicado de executar.



**TIP**

Para superar essas dificuldades, um número de fornecedores oferecem modelos de dados pré-treinado. Por exemplo, uma empresa pode fornecer centenas de milhares de imagens médicas pré-rotuladas para ajudar os clientes a criar um aplicativo que pode ajudar a imagens médicas de tela e detectar possíveis problemas de saúde.

## Reciclagem contínuo de Modelos

Atualmente, a maioria dos modelos de aprendizagem de máquina estiver offline. Estes modelos off-line são treinados usando dados treinados e depois implantado. Depois de um modelo off-line é implantado, o modelo subjacente não muda como ele é exposto a mais dados. O problema com modelos off-line é que eles presumem os dados de entrada permanecerá razoavelmente consistente.

Ao longo dos próximos anos, você vai ver mais modelos de aprendizagem de máquina disponível para uso. Como esses modelos são constantemente actualizados com novos dados, melhor os modelos estarão em análise preditiva. No entanto, as preferências e as tendências mudam, e modelos off-line não conseguem se adaptar com as mudanças de dados de entrada. Por exemplo, tomar a situação em que um modelo de aprendizagem de máquina faz previsões sobre a probabilidade de que um cliente vai produzir. O modelo poderia ter sido muito preciso quando foi implantado, mas como novos concorrentes, mais flexíveis surgir, e uma vez que os clientes têm mais opções, a sua probabilidade de churn vai aumentar. Porque o modelo original foi treinado em dados mais antigos antes de novos operadores no mercado surgiu, que deixará de dar à organização previsões precisas. Por outro lado, se o modelo está online e constante adaptação com base em dados de entrada,

## Aprendizado de Máquina como um serviço vai crescer

Como os modelos e algoritmos que suportam a aprendizagem de máquina maduro, você verá a crescente popularidade do Machine Learning como um Serviço (MLaaS). MLaaS descreve uma variedade de capacidades de aprendizado de máquina que são entregues através da nuvem. Fornecedores no MLaaS mercado oferta ferramentas como reconhecimento de imagem, reconhecimento de voz, visualização de dados e aprendizagem profunda. Um usuário normalmente upload de dados para a nuvem de um fornecedor e, em seguida, o cálculo de aprendizado de máquina é processado na nuvem.



**WARNING**

Alguns dos desafios de movimentação de grandes conjuntos de dados para a nuvem incluem custos de rede, compliance e riscos de governança e desempenho. No entanto, usando um serviço de nuvem, as organizações podem usar a aprendizagem de máquina sem o tempo inicial e os custos associados com a aquisição de hardware.

Além disso, MLaaS abstrai grande parte da complexidade envolvida com a aprendizagem de máquina. Por exemplo, uma equipe pode usar o processamento de linguagem natural (NLP) - uma ferramenta utilizada para interpretar o texto ou a imagem reconhecimento - para criar um diálogo entre humanos e máquinas. Ambos PNL e reconhecimento de imagem são adequados para a aplicação de serviços de nuvem que foi concebido para processar computação específica tarefas intensivas. As diferenças de desempenho são especialmente importantes quando o treinamento e iteração muitos modelos. Grande Graphic Processing Units (GPUs) são projetados para acelerar a renderização de imagens para que possam reduzir significativamente o tempo de ciclo.

## A maturação da PNL

Esperamos que na próxima década, a PNL vai amadurecer o suficiente para ser a norma para os usuários se comuniquem com sistemas através de uma interface escrita ou falada. PNL é a tecnologia que permite que as máquinas para compreender a estrutura eo significado das línguas faladas e escritas dos seres humanos. Além disso, a tecnologia de NLP permite que as máquinas informações de saída na linguagem falada compreendido por seres humanos. Os investigadores têm vindo a trabalhar em tecnologia de NLP por décadas, e aprendizado de máquina está ajudando a acelerar a implementação de sistemas de PLN. Atualmente, é muito difícil para as máquinas de compreender o contexto das palavras e frases. Através da aplicação de aprendizagem de máquina para a PNL, os sistemas são capazes de aprender o contexto e significado de palavras e frases. Tomemos por exemplo a frase "Um morcego voou em direção à multidão." A frase poderia estar se referindo a um taco de beisebol que um lançador inadvertidamente deixar de ir ou um mamífero voador que estava indo em direção a uma multidão de pessoas. Para entender o significado da frase, um sistema precisaria ingerir o contexto em torno dessa frase.

## Mais Automation vai agilizar Pipelines Machine Learning

Automatizando o processo de aprendizagem de máquina vai dar aos funcionários menos técnicos acesso a recursos de aprendizagem de máquina. Além disso, pela

acrescentando automação, usuários técnicos será capaz de se concentrar no trabalho mais desafiador ao invés de simplesmente automatizar tarefas repetitivas. Há muitos detalhes tediosos envolvidos com aprendizagem de máquina que são importantes, mas maduro para automação (por exemplo, limpeza de dados). visualização de dados é outra área onde a automação está ajudando a agilizar o processo de aprendizagem de máquina. Os sistemas podem ser projetados para selecionar a visualização mais adequado para um determinado conjunto de dados, tornando mais fácil para entender a relação entre pontos de dados.

## Hardware especializado irá melhorar o desempenho de Aprendizado de Máquina

Estamos nos aproximando de uma época onde hardware sofisticado é agora acessível. Por isso, muitas organizações podem adquirir hardware que é poderoso o suficiente para processar rapidamente algoritmos de aprendizagem de máquina. Além disso, este equipamento poderoso remove o gargalo de processamento de aprendizagem máquina, permitindo assim a aprendizagem de máquina a ser incorporado em mais aplicações.

Tradicionalmente, os CPUs foram utilizados para apoiar o processo de formação de aprendizagem profunda com resultados mistos. Estas CPUs são problemáticos por causa da maneira complicada que eles processam passos em uma rede neural. Em contraste, as GPUs têm centenas de núcleos mais simples que permitem que milhares de segmentos de hardware concorrentes. Devido à importância de GPUs em aplicações de aprendizagem de profundidade, tem havido considerável pesquisa vai para a tecnologia, a fim de oferecer processadores mais poderosos. Nuvem fornecedores de computação também reconhecem o valor de GPUs, e mais deles estão oferecendo ambientes GPU na nuvem.

Além de GPUs, os pesquisadores estão usando Campo-Programmable Gate Arrays (FPGAs) para executar com êxito cargas de trabalho de aprendizagem de máquina. Às vezes FPGAs superar GPUs ao executar rede neural e operações de aprendizagem profundas.

## Automatizar algoritmo de seleção e teste Algoritmos

cientistas de dados normalmente precisa entender como usar dezenas de algoritmos de aprendizado de máquina específicas. No Capítulo 3, discutimos



os principais tipos de algoritmos de aprendizado de máquina. Uma variedade de algoritmos são usados para diferentes tipos de dados ou diferentes tipos de perguntas que você está tentando responder.

Escolhendo o algoritmo certo para criar um modelo de aprendizagem de máquina nem sempre é fácil. Um cientista de dados pode experimentar vários algoritmos diferentes até encontrar aquele que cria o melhor modelo. Esse processo leva tempo e requer um alto grau de especialização. Automação está sendo aplicada para ajudar a acelerar a tarefa de seleção algoritmo. Usando automação, os cientistas de dados é capaz de concentrar-se rapidamente em apenas um ou dois algoritmos em vez de testar manualmente muitos mais. Além disso, essa automação ajuda os desenvolvedores e analistas com menos de aprendizagem de máquina experiência de trabalho com algoritmos de aprendizado de máquina.

## Transparência e Confiança-se uma exigência

Compreender não apenas como mas por um modelo de aprendizagem de máquina recomenda um resultado específico será essencial para confiar nos resultados. Um modelo de aprendizagem profunda utilizado para a digitalização de imagens médicas pode sinalizar uma imagem para um potencial de crescimento canceroso. No entanto, simplesmente identificar a imagem não é suficiente. O médico precisa entender por que o modelo de máquina pensou que o crescimento era canceroso. Que informações foram analisadas para levar o modelo para concluir o diagnóstico? O médico deve estar convencido de que os resultados são confirmados pelos dados.

## Aprendizado de Máquina como um processo End-to-End

Agora que estamos nos movendo para uma era de comercialização de aprendizagem de máquina, vamos começar a ver a aprendizagem de máquina como um processo end-to-end de uma perspectiva de desenvolvimento e operações. Isto significa que o processo inclui identificar os dados corretos para resolver um problema complexo, garantindo que os dados são devidamente treinados, modelado, e conseguiu em uma base contínua. Este ciclo de vida de aprendizagem de máquina é fundamental, porque há tanta coisa em jogo. modelos de aprendizagem de máquina pode ser uma ferramenta poderosa para prever o futuro.

# The future with machine learning

New competitors are emerging and customer expectations have never been higher. Your business and reputation requires you to adapt to new trends and anticipate customer demands. Machine-learning technology is embedded in applications throughout enterprises in order to improve performance, increase customer satisfaction, reduce customer churn, and boost revenue. In this book, you discover what machine learning is, how to adopt machine learning in your company, and how machine learning can help your company.

## Inside...

- What is machine learning?
- Explaining the business imperative
- The key machine learning algorithms
- Skills for your data science team
- How businesses use machine learning
- The future of machine learning



### Learn more at:

[IBM.com/machinelearning](http://IBM.com/machinelearning)  
[IBM.com/datascience](http://IBM.com/datascience)

**Judith Hurwitz**, President, Hurwitz & Associates, is a consultant and thought leader. **Daniel Kirsch**, principle analyst, Hurwitz & Associates, is a researcher and consultant in machine learning, cloud, and security.

Go to **Dummies.com**®  
for videos, step-by-step photos,  
how-to articles, or to shop!

**for  
dummies**®  
A Wiley Brand

ISBN: 978-1-119-45495-3  
Part #: IMM14209USEN-00  
Not for resale

## **END WILEY User License Agreement**

Ir para [www.wiley.com/go/eula](http://www.wiley.com/go/eula) acessar EULA ebook da Wiley.