

ANA RAQUEL

POSTECH

IA PARA DEVS

MACHINE LEARNING

AULA 01

SUMÁRIO

O QUE VEM POR AÍ?	3
HANDS ON	4
SAIBA MAIS	5
O QUE VOCÊ VIU NESTA AULA?	13
REFERÊNCIAS.....	14

EMSE

O QUE VEM POR AÍ?

Você sabe o que é o aprendizado de máquina? Basicamente, consiste em **ensinar o computador** a aprender com os dados e algoritmos matemáticos. Agora, convido você para embarcar nessa jornada da ciência de dados e aprender sobre os tipos de aprendizado de máquina e como podemos utilizá-los para gerar insights para os negócios. Vamos lá?!



HANDS ON

Você sabe como identificar um problema que pode ser resolvido com aprendizado de máquina? Para identificar e entender os tipos de aprendizado de máquina, vamos olhar para os dados e para as suas características, e quais são os principais conceitos que devemos levar em consideração na escolha do melhor algoritmo para solucionar um problema de negócio.



SAIBA MAIS

Aprendizado de máquina

Para explorarmos ainda mais o assunto aprendizado de máquina, trago aqui duas frases que podem resumir o que é o aprendizado de máquina:

- Aprendizado de máquina é o campo de estudo que dá aos computadores a habilidade de aprender sem ser explicitamente programado (Arthur Samuel, 1959) .
- Diz-se que um programa de computador aprende pela experiência **E** em relação a algum tipo de tarefa **T** e alguma medida de desempenho **P** se o seu desempenho em **T**, conforme medido por **P**, melhora com a experiência **E** (Tom Mitchell, 1997).

Ambas as frases de Samuel e Mitchell explicam muito bem o que é o aprendizado de máquina. Perceba que Samuel diz que a máquina pode aprender sem ser explicitamente programada (isso é possível, pois os algoritmos possuem inteligência para identificar comportamentos e particularidades nos dados com base em algumas funções matemáticas) e Mitchell descreve os passos do aprendizado de máquina (o aprendizado de máquina aprende com os dados alimentados em um modelo matemático e pode ser avaliado seu desempenho utilizando métricas estatísticas para melhorar os resultados e manter o modelo eficaz). Mas qual seria o benefício do uso de um aprendizado de máquina? O que uma empresa poderia estar ganhando em troca da construção de modelos matemáticos que são construídos no cenário de aprendizado de máquina?

Vamos utilizar o seguinte exemplo para você compreender melhor: suponha que um banco precise identificar quem são as boas pessoas pagadoras e as pessoas má pagadoras. Com um **volume de dados histórico** na qual são armazenadas **algumas informações** dos clientes (tal como histórico financeiro, renda e dados pessoais), seria possível identificar quem são as pessoas propensas a serem má pagadoras com base no **padrão aprendido pelo modelo matemático preditivo**. Observe a figura a seguir:

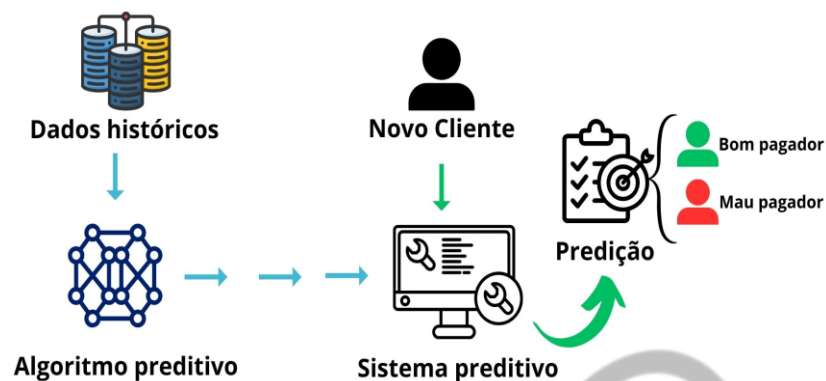


Figura 1 - Modelo preditivo
Fonte: elaborado pela autora (2024)

O algoritmo aprende o perfil das pessoas má pagadoras e boas pagadoras com base nas amostras de dados históricos, conseguindo, assim, distinguir seus comportamentos e aprendendo as particularidades com os dados. Com o comportamento dos padrões já aprendido, o modelo é capaz de generalizar novas entradas de dados (novos clientes nunca vistos antes pelo modelo), sendo possível prever se uma pessoa pode ser má ou boa pagadora.

Variáveis explicativas			Predição
Debitos Inadimplentes	Casa própria?	Carro próprio?	Condição de pagamento
100	Não	Não	Mau
1	Sim	Não	Bom
70	Não	Não	Mau

Figura 2 - Exemplo de base de dados para modelo preditivo
Fonte: elaborado pela autora (2024)

Obtendo essa “resposta” de antemão, é possível sugerir para a empresa meios alternativos para atender aos seus clientes, evitando assim, uma possível perda para a empresa.

Podemos pensar em vários cenários e assuntos para a criação de modelos preditivos no mercado:

- Finanças.
- Saúde.
- Educação.
- Agricultura.
- Logística.
- Marketing.
- Sustentabilidade.
- Varejo.
- Recursos Humanos.

Etapas do aprendizado de máquina

Quando estamos criando um algoritmo de Machine Learning, existem algumas etapas que são necessárias para construir o modelo preditivo. Observe a figura a seguir:

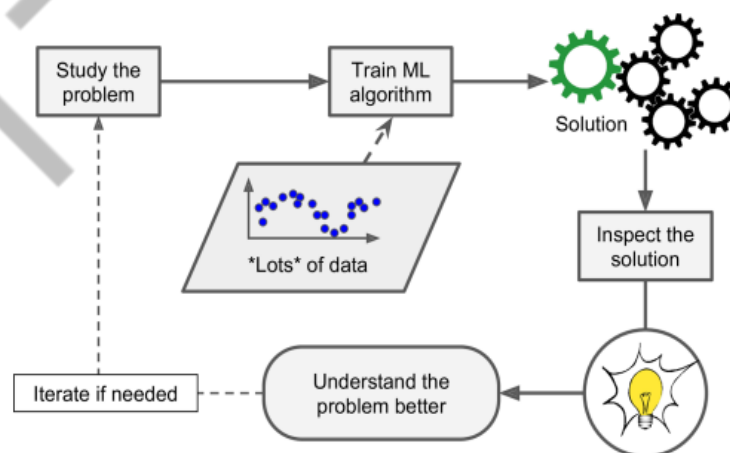


Figura 3 - Pipeline para modelo preditivo

Fonte: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2019)

Inicialmente, é necessário **estudar o problema de negócio** para começar a mapear e entender quais são as melhores features (características) que você precisa para **alimentar seu modelo preditivo**. Em seguida, é necessário escolher **qual é o tipo de algoritmo que melhor pode solucionar seu tipo de problema de negócio**. Com uma **grande amostra de dados**, **treinamos o modelo** para aprender os padrões e particularidades dos dados. Em seguida, é necessário **validar o desempenho do modelo** com uma **base de teste** validando de forma estatística a sua **eficácia**. Prontinho! Temos um modelo concluído para ser feito o deploy!

Caso o modelo não consiga generalizar os dados logo de primeira, um dos papéis fundamentais da pessoa cientista de dados é realizar a calibração deste modelo a fim de resolver o problema de **overfitting (quando o modelo decora os dados durante o treinamento)** e **underfitting (quando o modelo está muito fraco a fim de não aprender nenhum padrão com os dados)**.

Tipos de aprendizado de máquina

Dentro do mundo de Data Science, podemos citar três (3) principais tipos de aprendizado de máquina:

- ✓ Supervisionado.
- ✓ Não Supervisionado.
- ✓ Aprendizado por reforço.

Cada tipo de aprendizagem possui uma particularidade específica e dependendo do problema de negócio, podem ser utilizados em conjunto. Vamos conhecer mais de perto?

Aprendizado supervisionado

No aprendizado supervisionado, os dados de treinamento que você fornece ao algoritmo incluem as soluções desejadas, chamadas de **rótulos**. Por exemplo: imagine que você vai criar um modelo preditivo para identificar se um e-mail é spam ou não. Neste cenário, você já teria essa base de dados rotulada com a resposta que você procura, fazendo com que o algoritmo aprenda certo comportamento. Esse rótulo dentro da base é o que chamamos de **variável target** em um modelo preditivo supervisionado. Podemos entender nesse contexto, que essa coluna target ensina

nosso modelo que a partir de certas características (as variáveis explicativas), é possível prever uma classe alvo. Observe a figura a seguir:

Variáveis explicativas			variável alvo
Quantidade de Propagandas	Endereço de e-mail existente na agenda?	Quantidade de caracteres	
0	sim	1300	não
1	não	100	sim
0	sim	1000	não
2	não	900	sim
3	não	500	sim
0	sim	2000	não

Figura 4 - Exemplo base de dados para modelo supervisionado
Fonte: elaborado pela autora (2024)

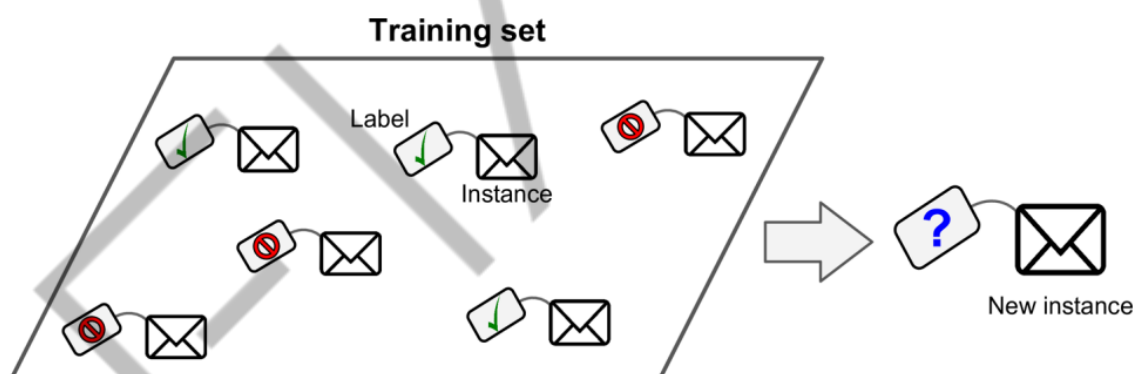


Figura 5 - Exemplo de modelo supervisionado
Fonte: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2019)

Quando o rótulo dos dados é um **valor numérico** e a base de dados possui um conjunto de características que descrevem esse valor, chamamos essa técnica de **“regressão”**.

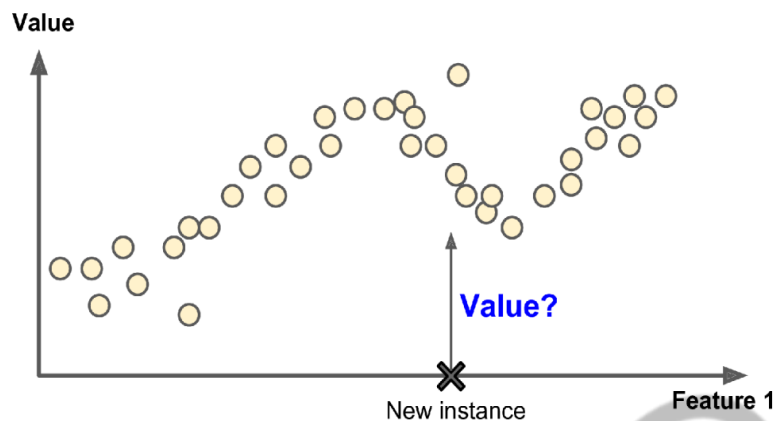


Figura 6 - Exemplo de modelo de regressão

Fonte: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2019)

Podemos citar alguns tipos de aprendizado supervisionado:

- K-Nearest Neighbors (KNN).
- Regressão Linear.
- Regressão Logística.
- Máquinas de Vetores de Suporte (SVM).
- Árvores de decisão.
- Árvores aleatórias.
- Redes Neurais.

Aprendizado não supervisionado

E quando não temos esses rótulos na base? No aprendizado não supervisionado, os dados de treinamento não são rotulados. Podemos dizer que o algoritmo aprende padrões sem um “professor”. Imagine aqui, por exemplo, uma amostra de dados na qual você quer criar grupos de clientes para encontrar perfis com base na similaridade de algum comportamento em comum entre eles. O aprendizado não supervisionado pode ter diferentes objetivos possíveis. Em alguns casos pode ser utilizado para criar uma regra preditiva na ausência de uma resposta rotulada. Os métodos de agrupamento podem ser utilizados para identificar grupos de dados significativos.

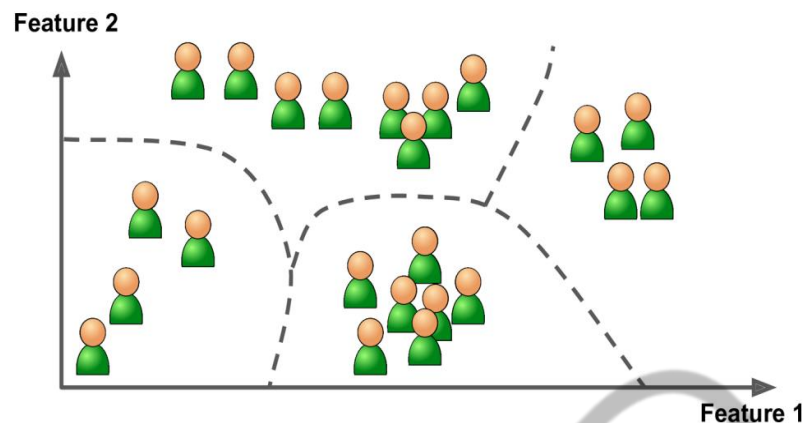


Figura 7 - Exemplo de aprendizagem não supervisionada
Fonte: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2019)

Podemos citar alguns tipos de modelos de aprendizagem não supervisionada tais como:

- K-Means.
- DBSCAN.
- Cluster hierárquico.
- Análise de componentes principais (PCA).
- T-distributed Stochastic Neighbor Embedding (T-SNE).
- Sistemas de recomendação.

Aprendizagem por reforço

Basicamente, funciona como tentativa e erro, que podem ser retribuídos com alguma recompensa quando o modelo de fato gera um acerto, e punição quando o modelo erra a previsão. O algoritmo observa a ação utilizando uma política e realiza a ação. Dependendo da escolha do algoritmo, o mesmo recebe uma recompensa ou uma punição. Ao receber a punição, o modelo atualiza a política de regras (etapa aprendida) e interagirá até encontrar uma boa política de regras.

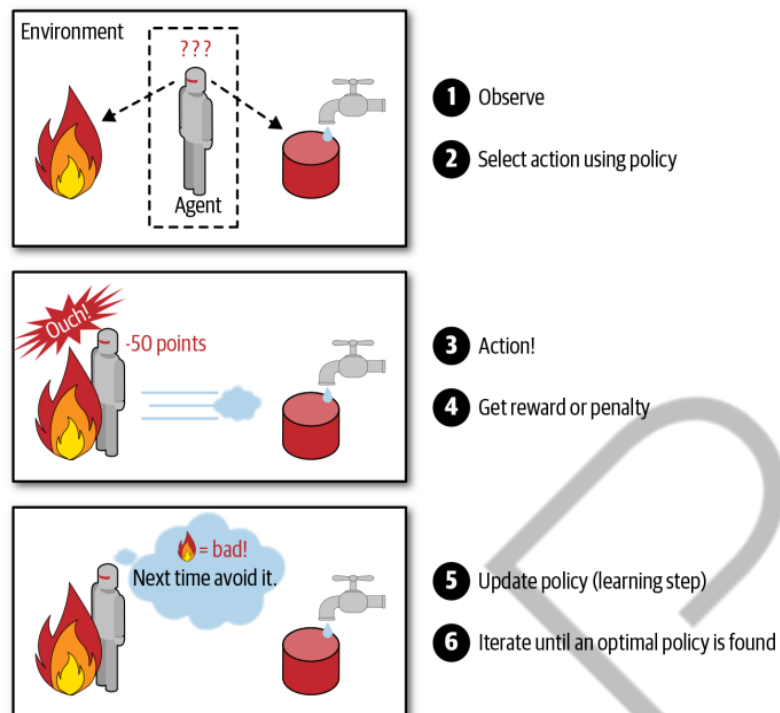


Figura 8 - Exemplo de aprendizagem por reforço
 Fonte: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2019)

O QUE VOCÊ VIU NESTA AULA?

Nessa aula, você viu uma introdução aos conceitos básicos e aplicações do aprendizado de máquina. Sua jornada na ciência de dados está apenas começando, então se prepare, pois nas próximas aulas você aprenderá como funciona essa “caixa mágica” do aprendizado de máquina.

EMSE

REFERÊNCIAS

GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow**. 2nd Edition. Massachusetts: O'Reilly Media, Inc. 2019.

EMASP

PALAVRAS-CHAVE

Aprendizagem Supervisionada, Aprendizagem Não Supervisionada, Aprendizagem por Reforço.

EMENDAS



POS TECH