

Machine Learning: Caso de Uso

Como a Zillow Transforma Dados
em Valor de Mercado

Fernando Gomes Cruz



Agenda

- Aprendizado de máquina
- Tipos Comuns de Aprendizado
- Aprendizado Supervisionado
- Caso de Uso: Zillow
- Solução: Neural Zestimate
- Exemplo
- Conclusões

O Que é Machine Learning?

É a ciência de ensinar computadores a aprender com dados para:

- Identificar padrões complexos.
- Fazer previsões precisas.
- Automatizar decisões.

Em vez de programar regras, nós "treinamos" um modelo.



Machine Learning

Uma Ideia Antiga, uma Revolução Recente

1950s: Conceitos nascem com Alan Turing e Arthur Samuel, que criou um jogo de damas que aprendia a jogar.

2010: A "tempestade perfeita" de Big Data e Poder Computacional (GPUs) torna o ML acessível e escalável.

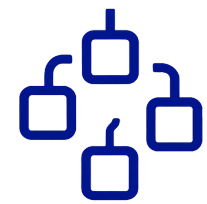
Hoje: É a tecnologia por trás de carros autônomos, assistentes virtuais e modelos de negócio inteiros.

Tipos Comuns de Aprendizado



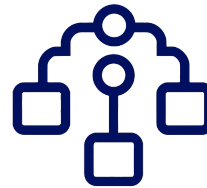
Aprendizado Supervisionado

Prever um resultado com base em exemplos passados (dados rotulados).



Aprendizado Não Supervisionado

Encontrar estruturas e grupos ocultos nos dados (ex: segmentação de clientes).



Aprendizado Semi-Supervisionado

Uma mistura dos dois, útil quando rotular dados é caro.



Aprendizado por Reforço

Aprender por tentativa e erro, com um sistema de recompensas (ex: treinar um robô ou um IA para jogos).

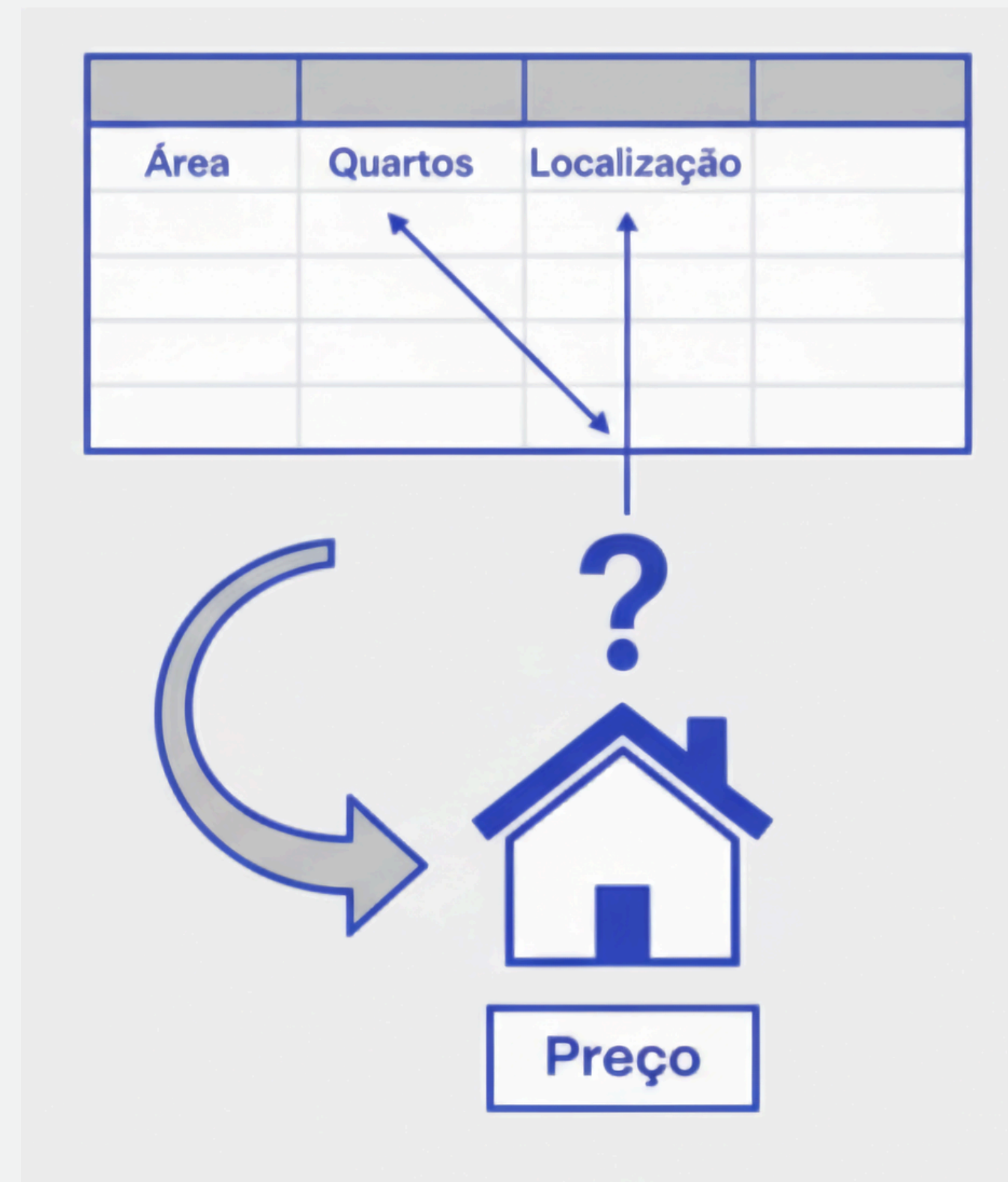
Aprendizado Supervisionado

Aprendendo com o Gabarito

O algoritmo aprende uma relação entre **Atributos (Features)** e um **Rótulo (Label)**.

- **Atributos:** As características de entrada (ex: área do imóvel, nº de quartos, localização).
- **Rótulo:** A resposta que queremos prever (ex: o preço final de venda).

Objetivo: Dado um novo conjunto de atributos, prever o rótulo com precisão.



■ ————— Caso de Uso



- **Quem:** Gigante de tecnologia imobiliária dos EUA.
- **Missão:** Trazer transparência e acesso ao mercado imobiliário.
- **Desafio Central:** Como estimar o valor de mais de 100 milhões de imóveis de forma rápida, precisa e automática?

A Solução: Neural Zestimate

É um modelo de Deep Learning Supervisionado que responde à pergunta:
"Quanto vale esta casa?"

The screenshot shows the Zillow website's 'Home Sale Proceeds Calculator'. The header includes navigation links: Buy, Rent, Sell, Get a mortgage, Find an Agent, the Zillow logo, Manage Rentals, Advertise, Help, and Sign In. The main heading is 'Home Sale Proceeds Calculator' with a subtext: 'Our home sale calculator estimates how much money you will make selling your home.' The calculator is divided into two main sections: 'Selling price & mortgage' and 'Estimated net proceeds'. The 'Selling price & mortgage' section includes a description: 'The amount you'd like to sell your home for and total remaining mortgage amount.' It has two input fields: 'Desired selling price *' with a value of '\$ 302,000' and 'Remaining mortgage owed' with a value of '\$ 0'. The 'Estimated net proceeds' section shows a large green result of '\$269,830'. Below this, a table lists the components: 'Desired selling price' (\$302,000), 'Remaining mortgage owed' (\$0), and 'Est. selling costs (10.65%)' (\$32,170). At the bottom, there is a section for 'Estimated prep & repair costs'.

Selling price & mortgage		Estimated net proceeds	
The amount you'd like to sell your home for and total remaining mortgage amount.		\$269,830	
Desired selling price *		Desired selling price	\$302,000
<input type="text" value="\$ 302,000"/>		Remaining mortgage owed	\$0
Remaining mortgage owed		Est. selling costs (10.65%)	\$32,170
<input type="text" value="\$ 0"/>			
Estimated prep & repair costs			

- **Modelo:** Rede Neural que processa uma quantidade massiva de dados.
- **Precisão:** Erro mediano de ~7.5% para casas fora do mercado. Uma melhoria de mais de 15% sobre modelos antigos.
- **Impacto:** É a ferramenta que atrai milhões de usuários e alimenta todo o ecossistema da Zillow.

Ingredientes do Modelo Supervisionado

O que a Zillow usa como **Atributos (Features)** para treinar seu modelo?



Dados do Imóvel: Área, nº de quartos/banheiros, ano de construção, etc.



Dados Públicos: Impostos, registros de propriedade.



Transações Históricas: Preços de venda de casas similares na mesma região.



Fatores Geográficos: Dados de localização em múltiplas escalas (bairro, cidade, estado).



Fatores Temporais: Sazonalidade do mercado, tendências de valorização ao longo do tempo.

O **Rótulo (Label)** é sempre o preço de venda real de uma transação concluída.

Exemplo



Aplicação de modelos de aprendizado supervisionado no dataset **California Housing Dataset** do scikit-learn.

Dados relativos ao censo de 1990.

<https://github.com/neocrz/housing-price-predictor>

Previsor de Preços de Imóveis e Comparador de Modelos

Selecione o Modelo:

Gradient Boosting

Renda Média (x\$10k USD)

8,5

Idade do Imóvel (anos)

15

Média de Cômodos

7

Média de Dormitórios

1,1

População (do setor)

950

Ocupação Média (moradores)

2,6

Latitude

34,05

Longitude

-118,4

Prever Preço Provável

\$469,086.05

Previsor de Preços de Imóveis e Comparador de Modelos

Selecione o Modelo:

Random Forest



Renda Média (x\$10k USD)

4,2



Idade do Imóvel (anos)

25



Média de Cômodos

5,5



Média de Dormitórios

1,05



População (do setor)

1400



Ocupação Média (moradores)

3,1



Latitude

38,5



Longitude

-121,5



Prever Preço Provável

\$150,643.00

Conclusão

O Aprendizado Supervisionado é uma das ferramentas mais poderosas para transformar dados brutos em **decisões estratégicas e valor de negócio tangível**.

O caso da Zillow mostra que, com o problema certo e os dados corretos, um modelo de ML pode se tornar o coração de uma empresa multibilionária.