

A CIÊNCIA ALÉM DOS DADOS

Estamos cada vez ganhando mais conhecimento e mais “**poder**” para utilizar os dados de acordo com o que desejamos.

ÉTICA

Ética é o ramo da filosofia que se dedica ao estudo dos valores morais e dos princípios que guiam o comportamento humano. Ela busca responder questões sobre o que é certo e o que é errado, o que é bom e o que é mau, bem como investigar os fundamentos e justificações para essas distinções. A ética lida com questões fundamentais sobre como devemos agir individual e coletivamente em diferentes situações da vida. Ela procura estabelecer padrões de conduta que promovam o bem-estar, a justiça, a equidade e o respeito mútuo entre as pessoas.

MORAL

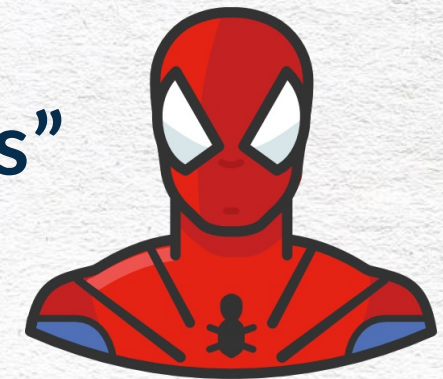
conjunto de valores, individuais ou coletivos, considerados universalmente como norteadores das relações sociais e da conduta dos homens.

Definição

A CIÊNCIA ALÉM DOS DADOS

Estamos cada vez ganhando mais conhecimento e mais “**poder**” para utilizar os dados de acordo com o que desejamos.

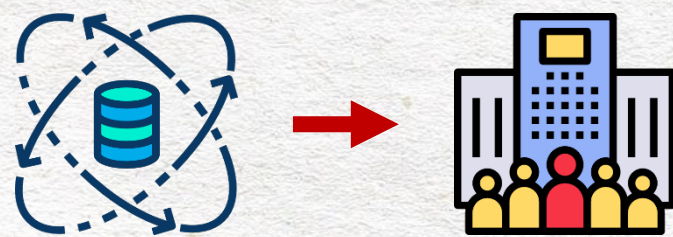
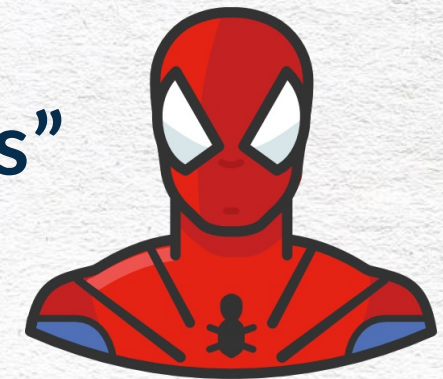
“Com grandes poderes, vêm grandes responsabilidades”



A CIÊNCIA ALÉM DOS DADOS

Estamos cada vez ganhando mais conhecimento e mais “**poder**” para utilizar os dados de acordo com o que desejamos.

“Com grandes poderes, vêm grandes responsabilidades”

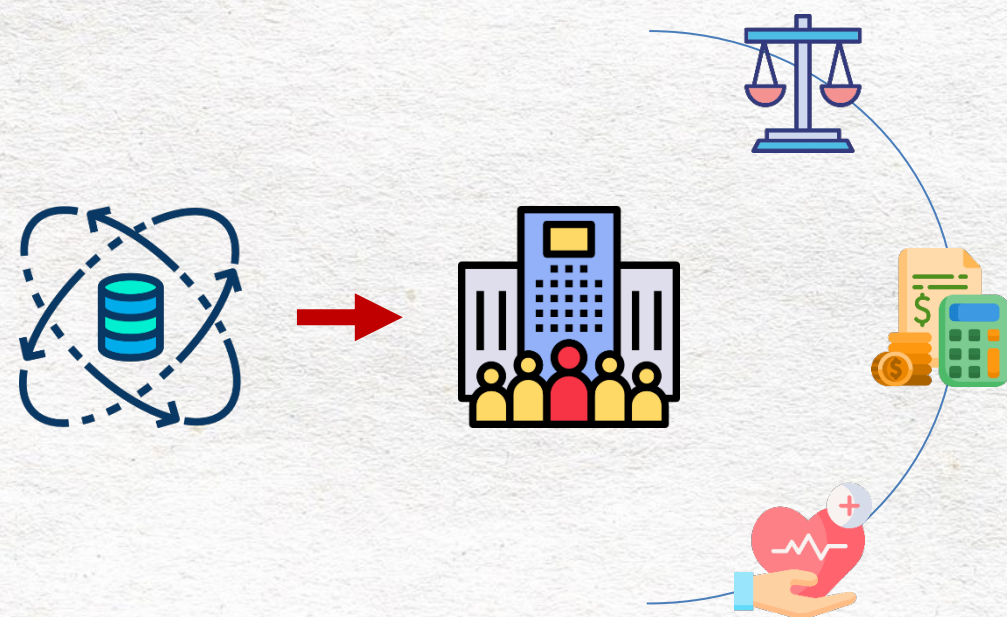


Se não tomarmos cuidado, podemos ficar tão focados em resolver o problema do negócio...

A CIÊNCIA ALÉM DOS DADOS

Estamos cada vez ganhando mais conhecimento e mais “**poder**” para utilizar os dados de acordo com o que desejamos.

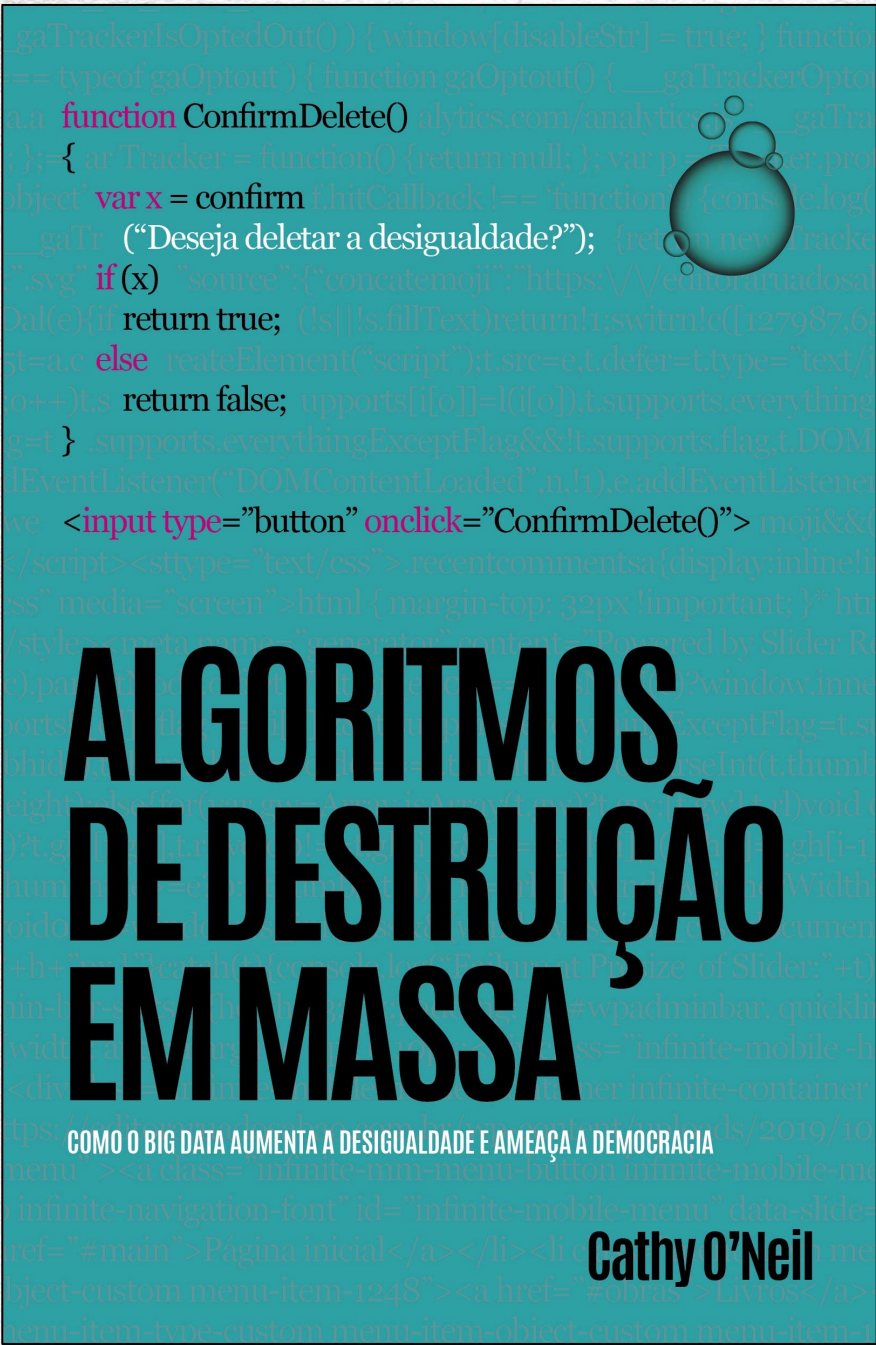
“Com grandes poderes, vêm grandes responsabilidades”



Se não tomarmos cuidado, podemos ficar tão focados em resolver o problema do negócio...

...que acabamos esquecendo de garantir que o nosso modelo esteja realmente sendo bom para a sociedade

A CIÊNCIA ALÉM DOS DADOS



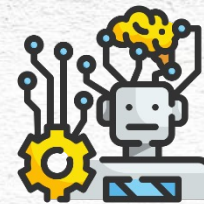
Algoritmos de destruição em massa
COMO O BIG DATA AUMENTA A DESIGUALDADE E AMEAÇA A DEMOCRACIA
Cathy O`Neil

A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

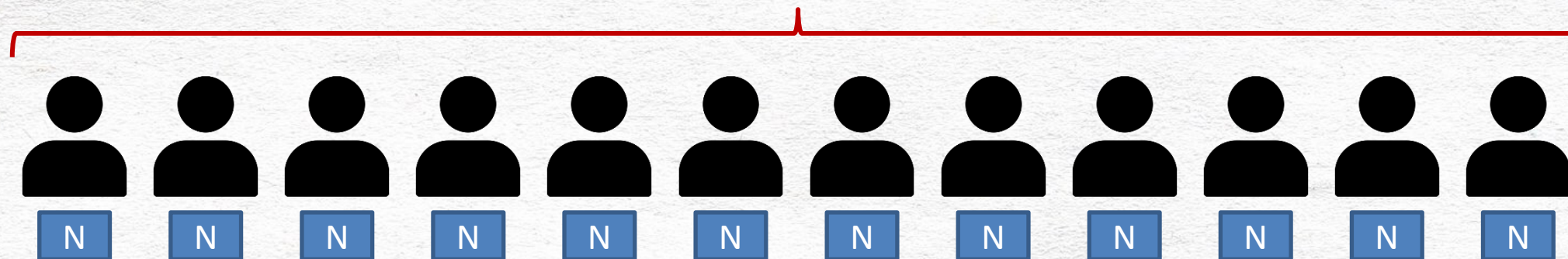
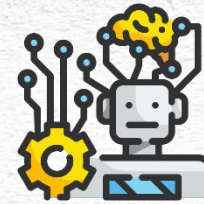


A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

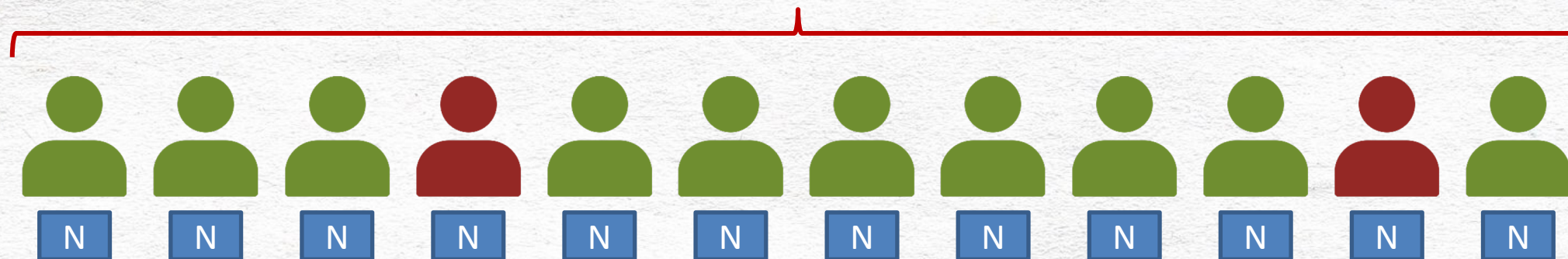
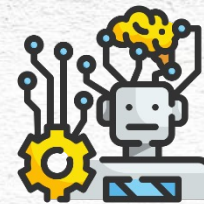


A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

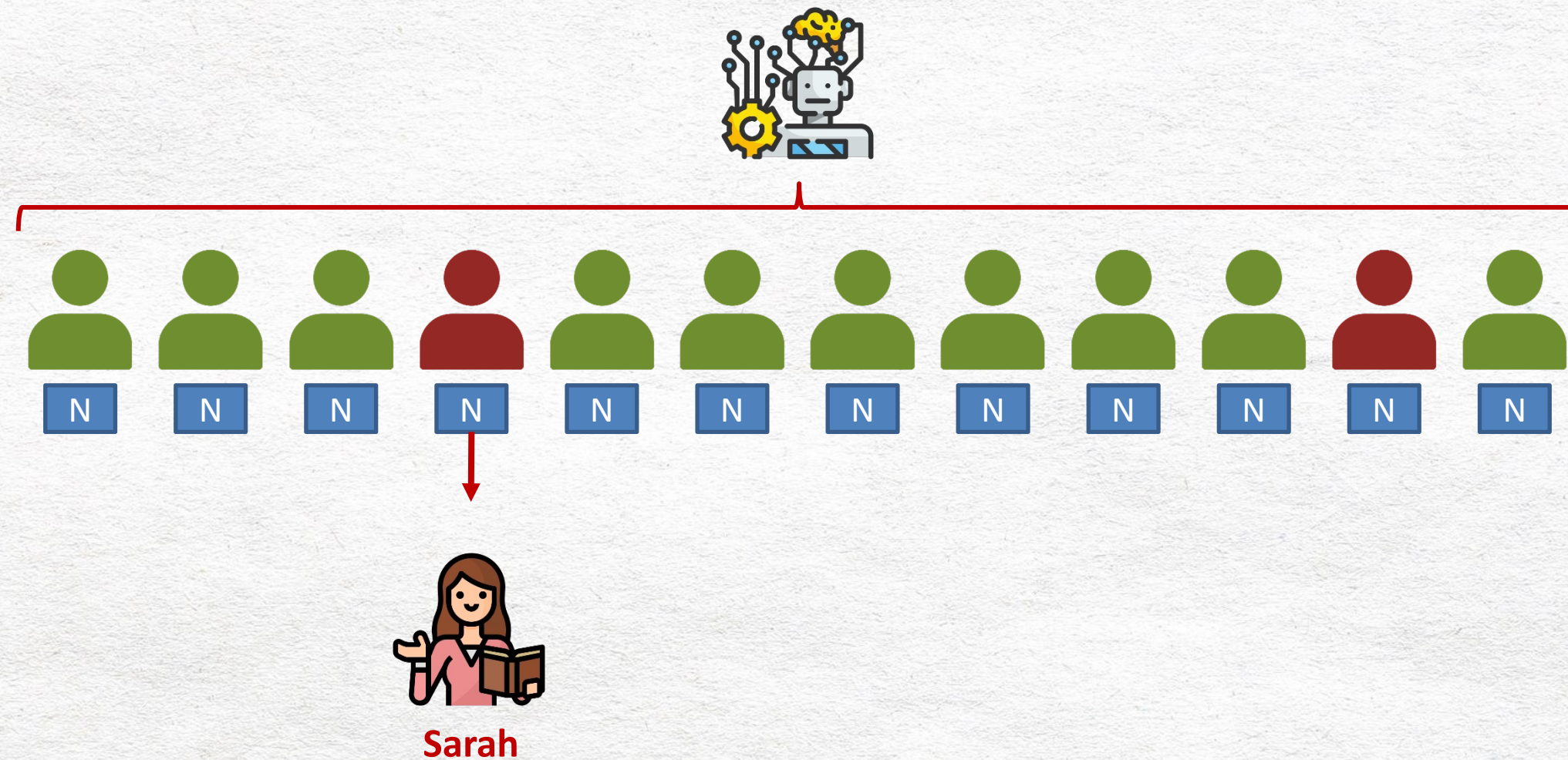


A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

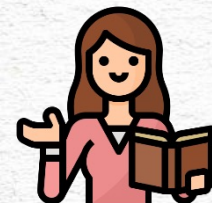


A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)



Sarah

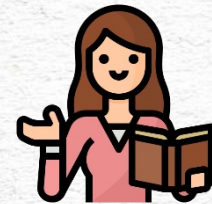
A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

Alunos gostavam dela
Tinha uma boa avaliação dos pais
Feedbacks positivos dos diretores



Sarah

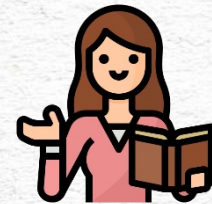
A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

Alunos gostavam dela
Tinha uma boa avaliação dos pais
Feedbacks positivos dos diretores



Sarah

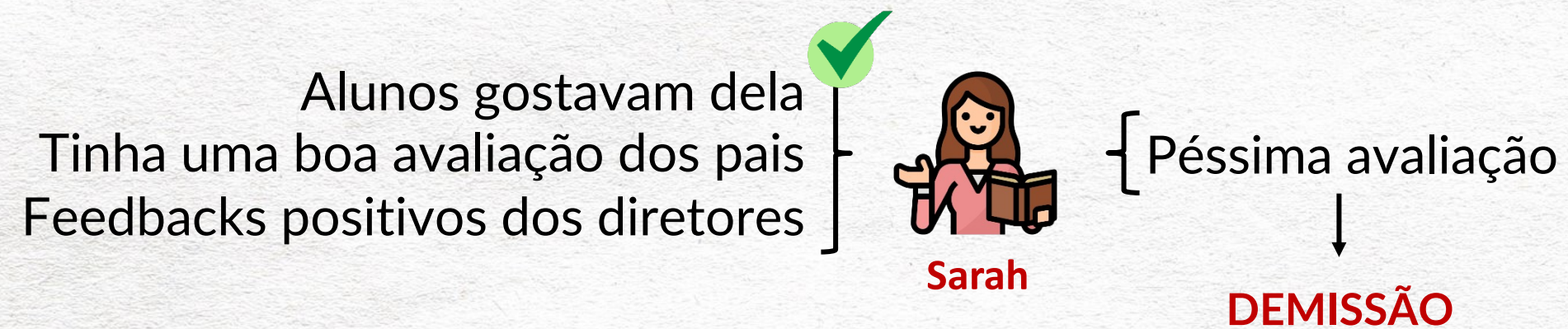
{ Péssima avaliação

A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

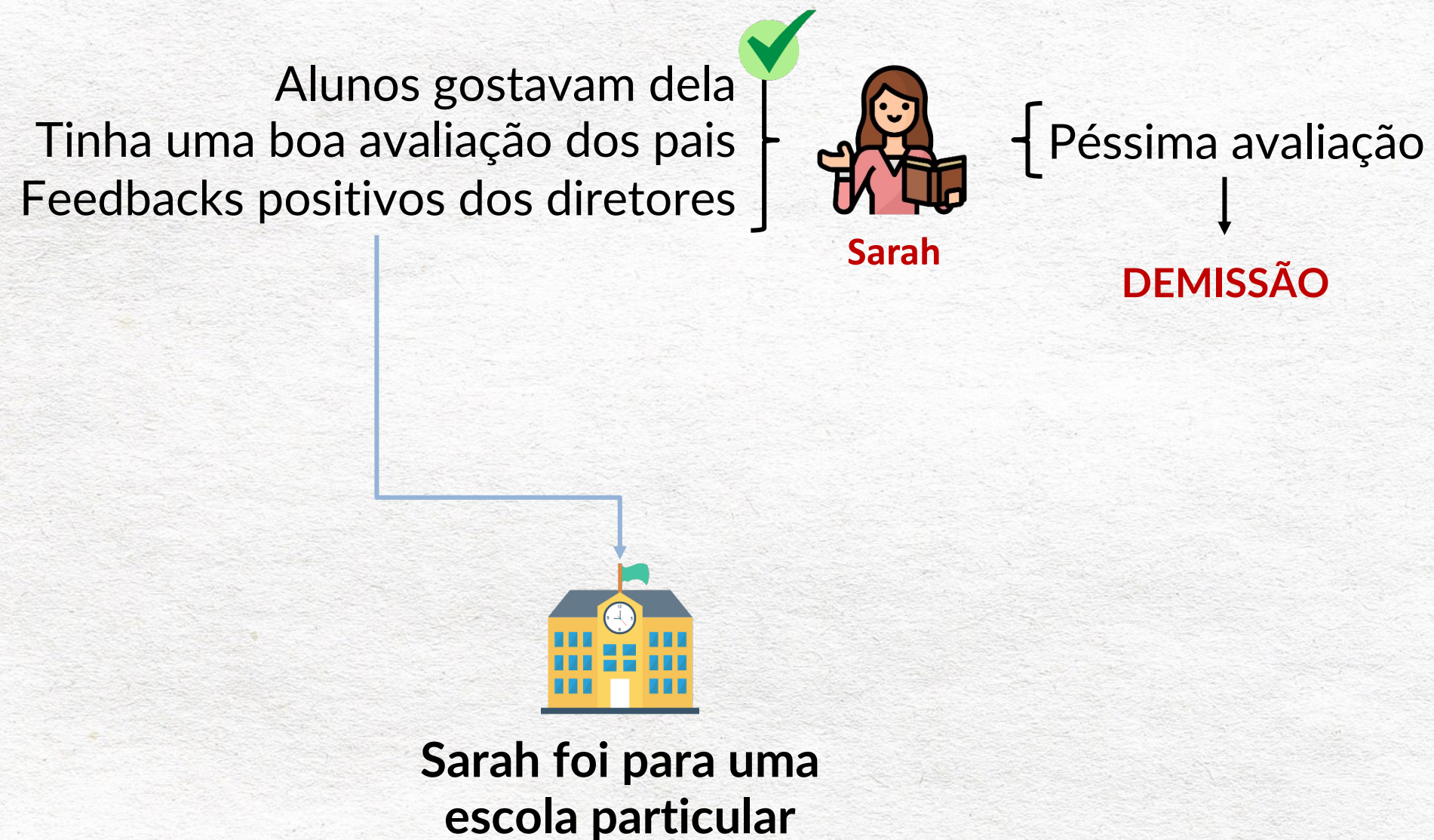


A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

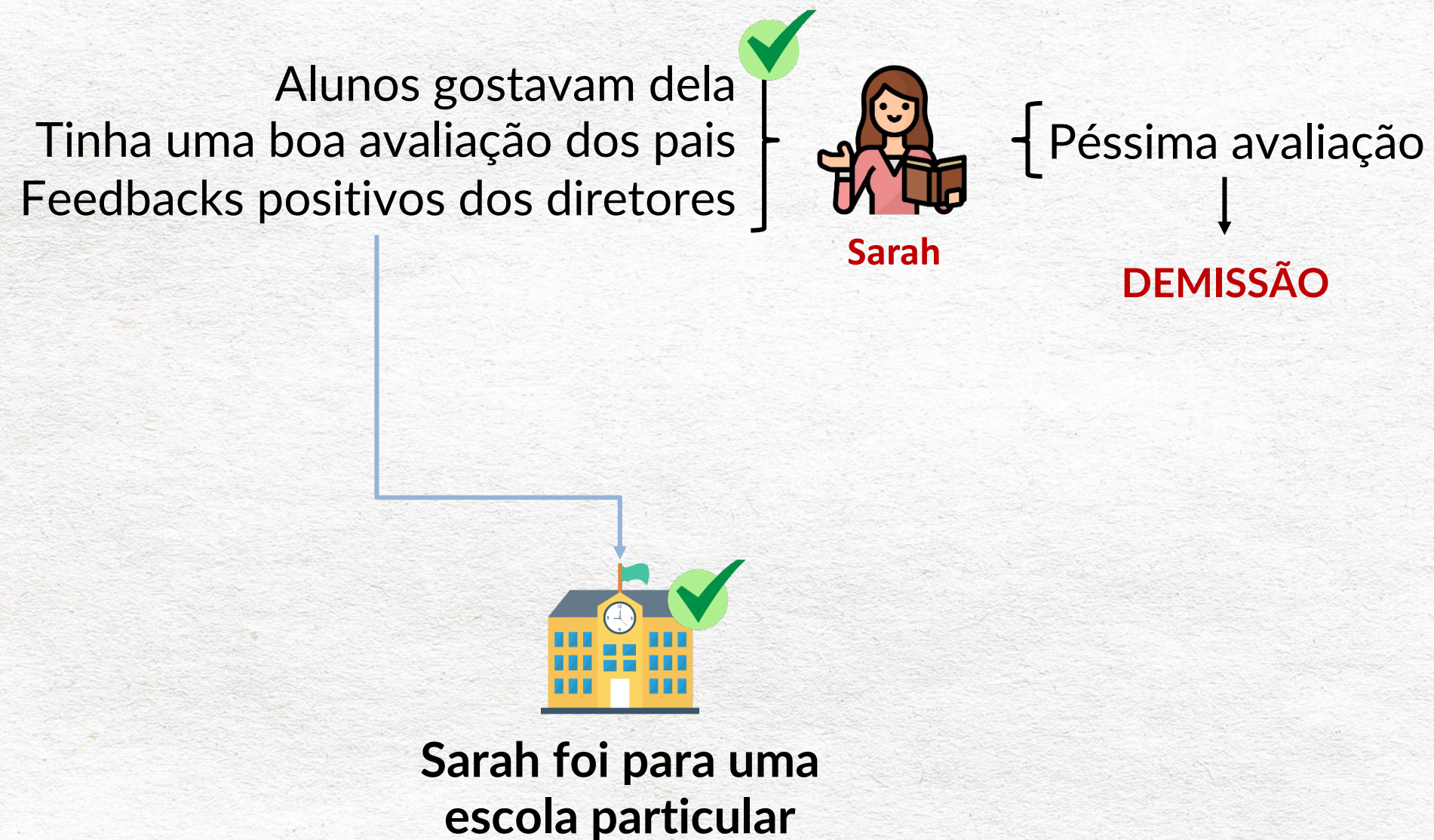


A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

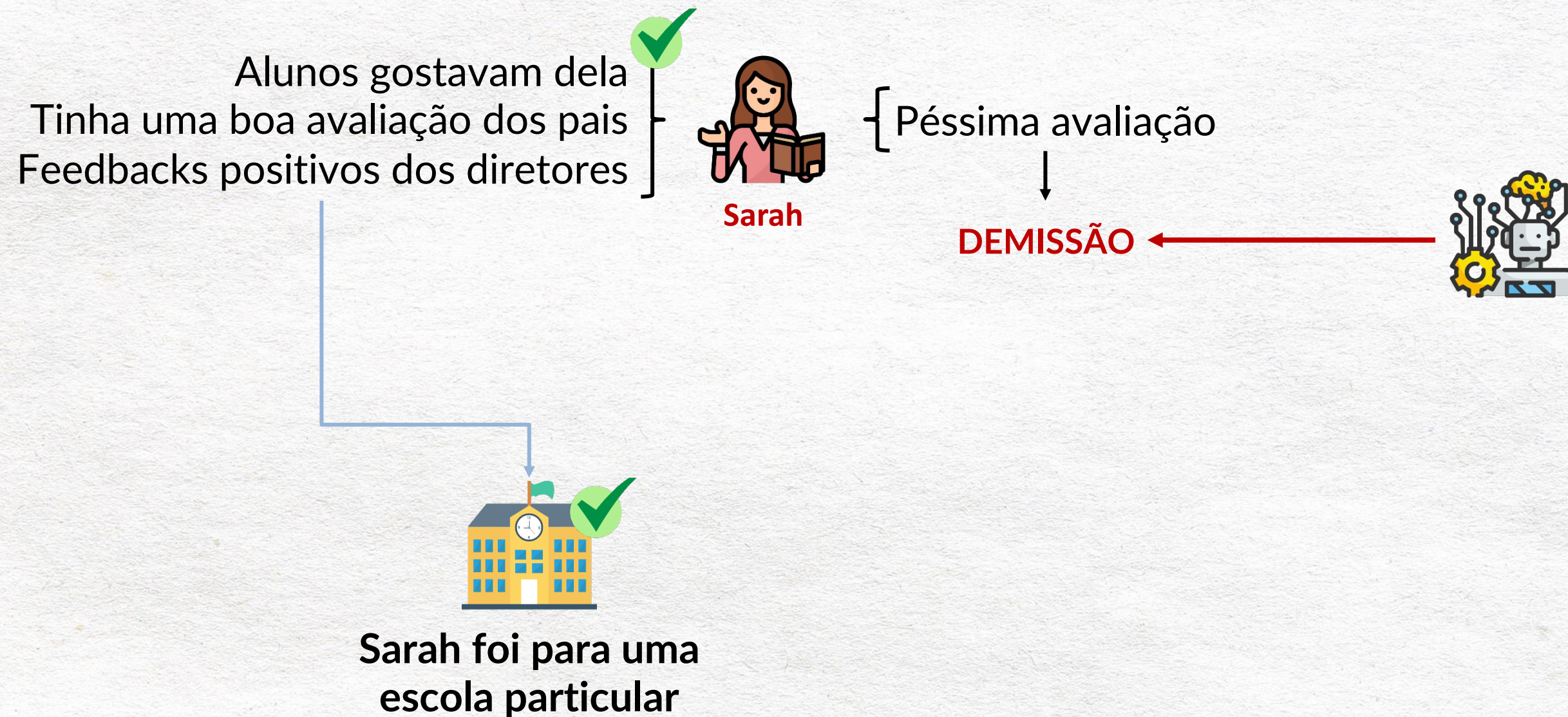


A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

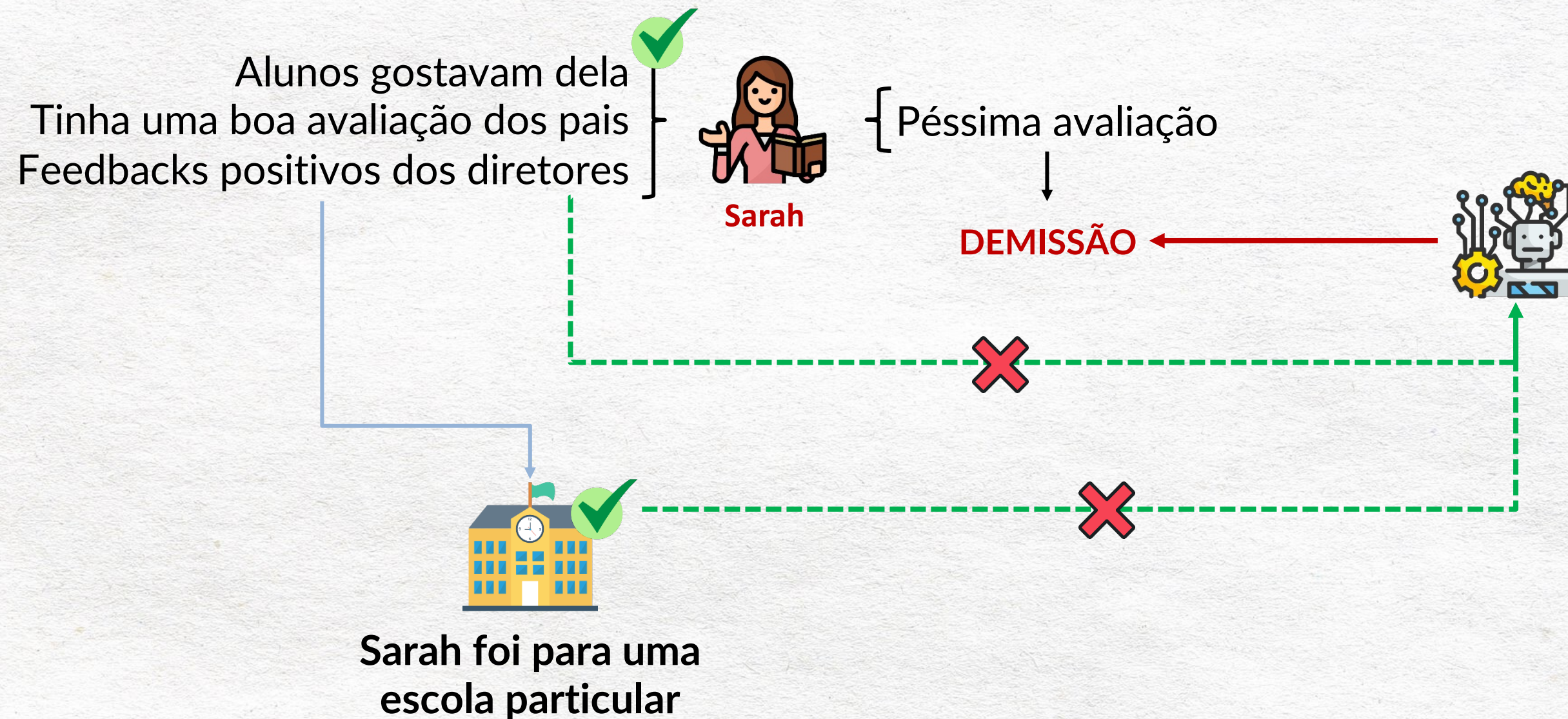


A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

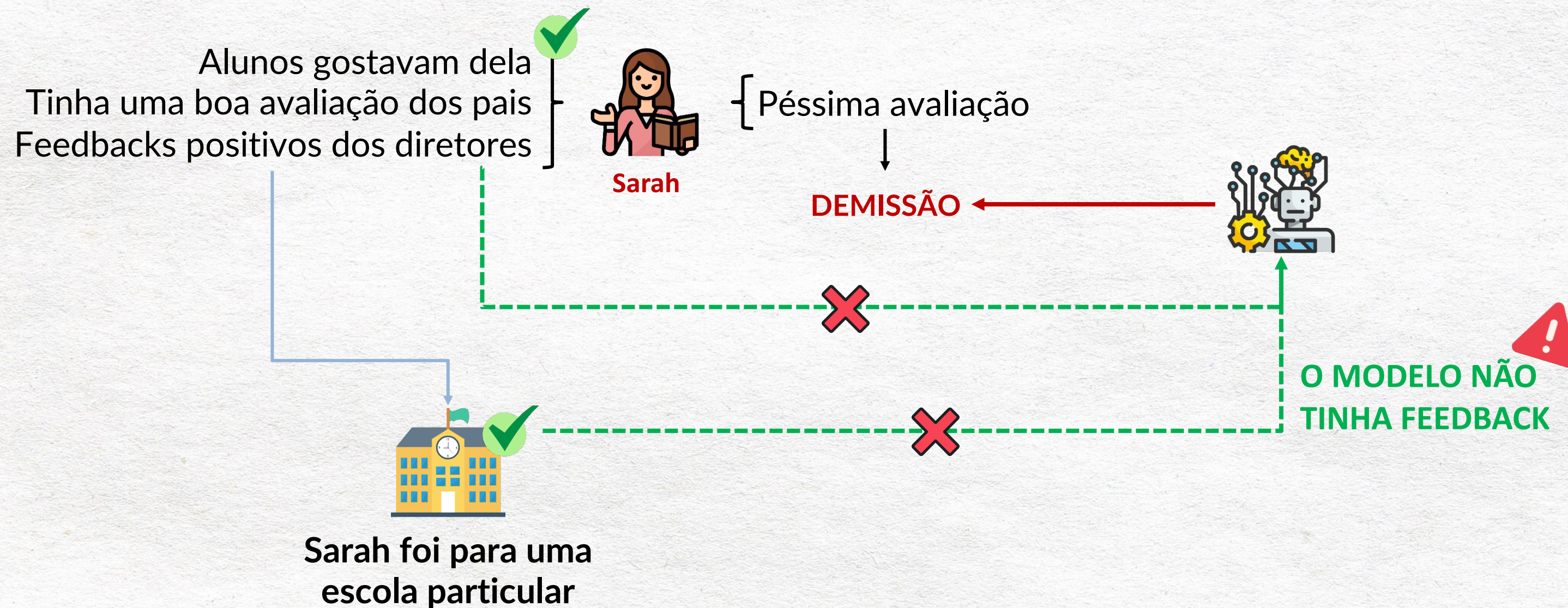


A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)

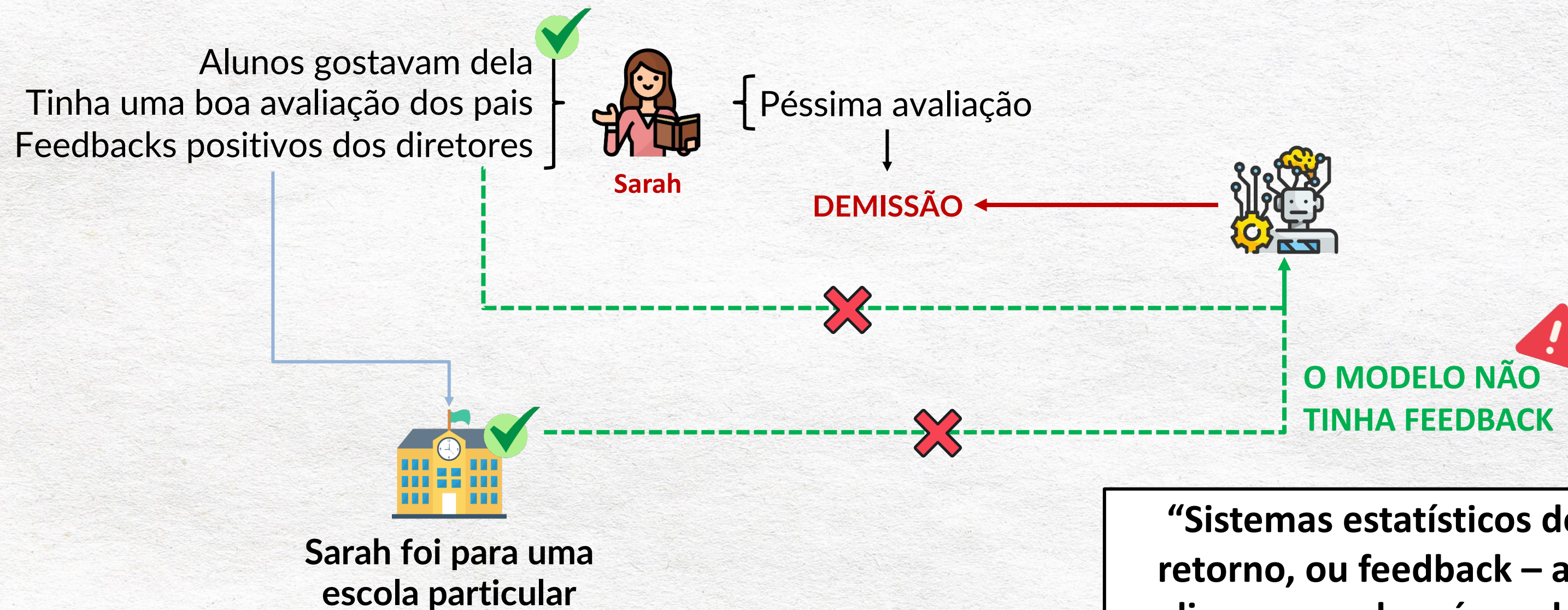


A CIÊNCIA ALÉM DOS DADOS



- Melhorar o desempenho das escolas:

- Prefeito de Washington DC em 2007 criou um projeto
- 2009: implementado o **IMPACT** (ferramenta de avaliação de professores)



“Sistemas estatísticos demandam retorno, ou feedback – algo que os digam quando saíram dos trilhos”

A CIÊNCIA ALÉM DOS DADOS



- Testes de personalidade:

- Não tem feedback, são obscuros para o usuário

“O computador aprendeu com os humanos a discriminar, e realizou esse trabalho com uma eficiência de tirar o fogo”

A CIÊNCIA ALÉM DOS DADOS



- Testes de personalidade:

- Não tem feedback, são obscuros para o usuário

“O computador aprendeu com os humanos a discriminar, e realizou esse trabalho com uma eficiência de tirar o fogo”



A Xerox **removeu a distância ao trabalho como critério** da modelagem pois percebeu que isso era discriminatório com bairros pobres

A CIÊNCIA ALÉM DOS DADOS



- Testes de personalidade:

- Não tem feedback, são obscuros para o usuário

“O computador aprendeu com os humanos a discriminar, e realizou esse trabalho com uma eficiência de tirar o fogo”



A Xerox **removeu a distância ao trabalho como critério** da modelagem pois percebeu que isso era discriminatório com bairros pobres

Reduziu a eficiência mas aumentou a justiça com os candidatos, além de deixar de excluir bons candidatos apenas por morarem longe

A CIÊNCIA ALÉM DOS DADOS



- Testes de personalidade:

- Não tem feedback, são obscuros para o usuário

“O computador aprendeu com os humanos a discriminar, e realizou esse trabalho com uma eficiência de tirar o fogo”



A Xerox **removeu a distância ao trabalho como critério** da modelagem pois percebeu que isso era discriminatório com bairros pobres

Reduziu a eficiência mas aumentou a justiça com os candidatos, além de deixar de excluir bons candidatos apenas por morarem longe



Porém sabemos que muitas empresas ainda usam nome, endereço, localização geográfica, score de crédito, etc

A CIÊNCIA ALÉM DOS DADOS



- Eleições:



1968: Nixon usou grupos de foco para diferentes regiões e demografias

- Separou o público e enviou via mala direta o que mais poderia influenciar aquele eleitor a votar nele

A CIÊNCIA ALÉM DOS DADOS



- Eleições:



1968: Nixon usou grupos de foco para diferentes regiões e demografias

- Separou o público e enviou via mala direta o que mais poderia influenciar aquele eleitor a votar nele



~2010: Facebook criou uma campanha para **encorajar as pessoas a votarem**

- **FERRAMENTA: “megafone do eleitor”**

- Começaram a testar apresentar para um grupo de pessoas 6 amigos que haviam clicado em “Eu votei” de forma aleatória para ver se isso o influenciaria a também clicar em “Eu votei”

A CIÊNCIA ALÉM DOS DADOS



- Eleições:



1968: Nixon usou grupos de foco para diferentes regiões e demografias

- Separou o público e enviou via mala direta o que mais poderia influenciar aquele eleitor a votar nele



~2010: Facebook criou uma campanha para **encorajar as pessoas a votarem**

- **FERRAMENTA: “megafone do eleitor”**

- Começaram a testar apresentar para um grupo de pessoas 6 amigos que haviam clicado em “Eu votei” de forma aleatória para ver se isso o influenciaria a também clicar em “Eu votei”

- 2 anos depois:

- Receber mais notícias postadas pelos amigos que outros tipos de posts
- Aumentou de 64 para 67% da participação nas eleições

A CIÊNCIA ALÉM DOS DADOS



- Eleições:



1968: Nixon usou grupos de foco para diferentes regiões e demografias

- Separou o público e enviou via mala direta o que mais poderia influenciar aquele eleitor a votar nele



~2010: Facebook criou uma campanha para **encorajar as pessoas a votarem**

- **FERRAMENTA: “megafone do eleitor”**

- Começaram a testar apresentar para um grupo de pessoas 6 amigos que haviam clicado em “Eu votei” de forma aleatória para ver se isso o influenciaria a também clicar em “Eu votei”

- 2 anos depois:

- Receber mais notícias postadas pelos amigos que outros tipos de posts
 - Aumentou de 64 para 67% da participação nas eleições

“Quando seus amigos te entregam o jornal, coisas interessantes acontecem”
- Lada Adamic, cientista social computacional do Facebook

A CIÊNCIA ALÉM DOS DADOS



Precisamos estar cientes que **sempre que criamos um modelo estamos selecionando quais dados serão considerados**, então nós mesmos determinamos o que é importante de acordo com nossos objetivos e julgamentos



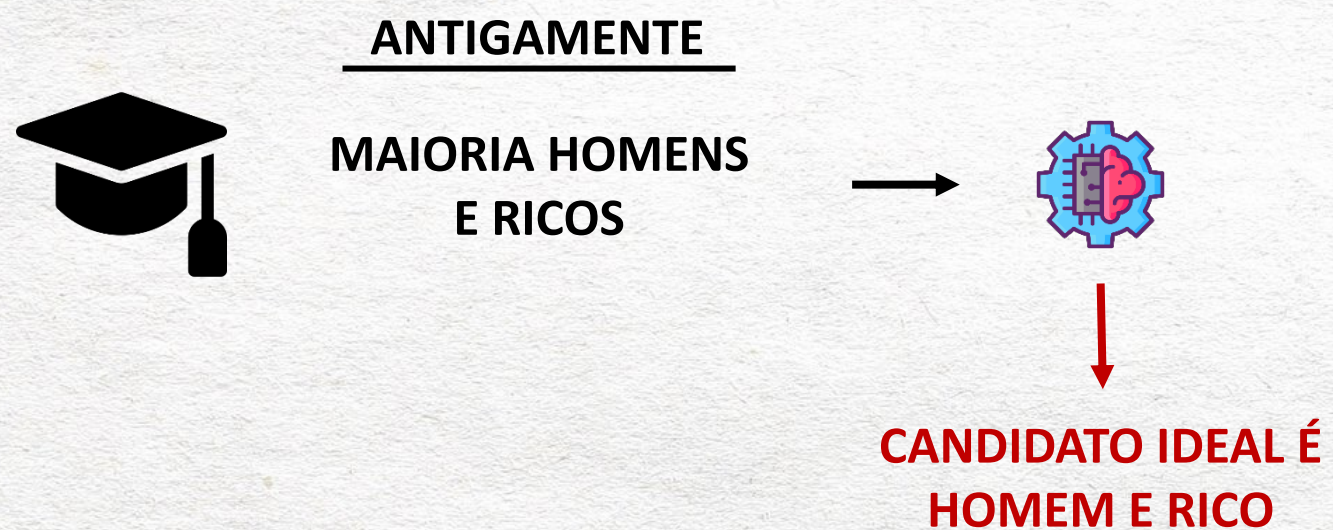
ANTIGAMENTE

MAIORIA HOMENS
E RICOS

A CIÊNCIA ALÉM DOS DADOS



Precisamos estar cientes que **sempre que criamos um modelo estamos selecionando quais dados serão considerados**, então nós mesmos determinamos o que é importante de acordo com nossos objetivos e julgamentos



A CIÊNCIA ALÉM DOS DADOS



Precisamos estar cientes que **sempre que criamos um modelo estamos selecionando quais dados serão considerados**, então nós mesmos determinamos o que é importante de acordo com nossos objetivos e julgamentos



A CIÊNCIA ALÉM DOS DADOS



Precisamos estar cientes que **sempre que criamos um modelo estamos selecionando quais dados serão considerados**, então nós mesmos determinamos o que é importante de acordo com nossos objetivos e julgamentos



A CIÊNCIA ALÉM DOS DADOS



Precisamos estar cientes que **sempre que criamos um modelo estamos selecionando quais dados serão considerados**, então nós mesmos determinamos o que é importante de acordo com nossos objetivos e julgamentos



DICA: modelos podem ser simples
- Detecção de incêndio: presença de fumaça



A CIÊNCIA ALÉM DOS DADOS



Precisamos estar cientes que **sempre que criamos um modelo estamos selecionando quais dados serão considerados**, então nós mesmos determinamos o que é importante de acordo com nossos objetivos e julgamentos



DICA: modelos podem ser simples
- Detecção de incêndio: presença de fumaça



A CIÊNCIA ALÉM DOS DADOS



Precisamos estar cientes que **sempre que criamos um modelo estamos selecionando quais dados serão considerados**, então nós mesmos determinamos o que é importante de acordo com nossos objetivos e julgamentos



DICA: modelos podem ser simples
- Detecção de incêndio: presença de fumaça



“Os pontos cegos de um modelo refletem o julgamento e prioridades dos seus criadores”

A CIÊNCIA ALÉM DOS DADOS



Sem o devido cuidado, as empresas podem tomar decisões para se ajustar aos **modelos** que vão contra o que é melhor para toda a sociedade!



U.S. News criou o
RANKING UNIVERSITÁRIO

A CIÊNCIA ALÉM DOS DADOS



Sem o devido cuidado, as empresas podem tomar decisões para se ajustar aos **modelos** que vão contra o que é melhor para toda a sociedade!



U.S. News criou o
RANKING UNIVERSITÁRIO

MÉTRICA: **considerava** o número de
ex alunos que faziam doações

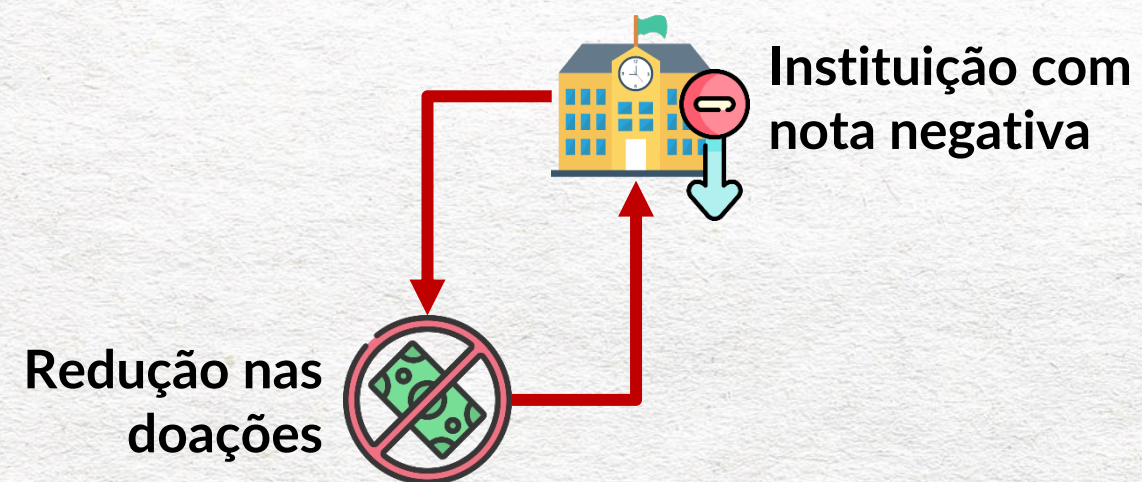
A CIÊNCIA ALÉM DOS DADOS

 Sem o devido cuidado, as empresas podem tomar decisões para se ajustar aos **modelos** que vão contra o que é melhor para toda a sociedade!



U.S. News criou o
RANKING UNIVERSITÁRIO

MÉTRICA: **considerava** o número de
ex alunos que faziam doações




A CIÊNCIA ALÉM DOS DADOS

 Sem o devido cuidado, as empresas podem tomar decisões para se ajustar aos modelos que vão contra o que é melhor para toda a sociedade!




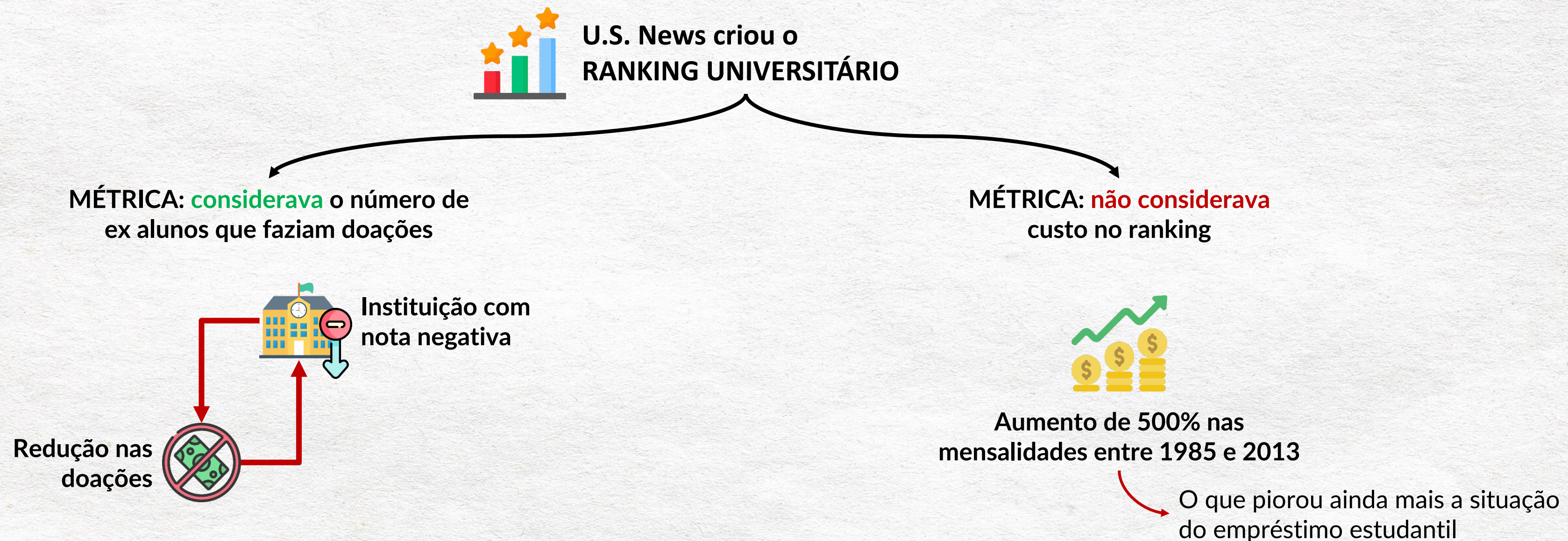
A CIÊNCIA ALÉM DOS DADOS

 Sem o devido cuidado, as empresas podem tomar decisões para se ajustar aos modelos que vão contra o que é melhor para toda a sociedade!



A CIÊNCIA ALÉM DOS DADOS

 Sem o devido cuidado, as empresas podem tomar decisões para se ajustar aos modelos que vão contra o que é melhor para toda a sociedade!



A CIÊNCIA ALÉM DOS DADOS



Precisamos estar sempre atentos!

`sklearn.datasets.load_boston`

```
sklearn.datasets.load_boston ( *, return_X_y = False )
```

[\[fonte\]](#)

DEPRECATED: `load_boston` está obsoleto na versão 1.0 e será removido na versão 1.2.

O conjunto de dados de preços imobiliários de Boston tem um problema ético. Você pode consultar a documentação desta função para obter mais detalhes.

Os mantenedores do scikit-learn, portanto, desencorajam fortemente o uso desse conjunto de dados, a menos que o objetivo do código seja estudar e educar sobre questões éticas em ciência de dados e aprendizado de máquina.

A CIÊNCIA ALÉM DOS DADOS



Precisamos estar sempre atentos!

`sklearn.datasets.load_boston`

```
sklearn.datasets.load_boston(*, return_X_y=False)
```

DEPRECATED: `load_boston` está obsoleto na versão 1.0 e será removido na versão 1.2.

O conjunto de dados de preços imobiliários de Boston tem um problema ético. Você pode consultar a documentação da função para obter mais detalhes.

Os mantenedores do scikit-learn, portanto, desencorajam fortemente o uso desse conjunto de dados. O código seja estudar e educar sobre questões éticas em ciência de dados e aprendizado de máquina.

7.1.1. Conjunto de dados de preços de casas em Boston

Características do conjunto de dados:

Número de Instâncias:	506
Número de Atributos:	13 preditivo numérico/categórico. O valor mediano (atributo 14) geralmente é o alvo.
Informações de atributo (em ordem):	<ul style="list-style-type: none">• CRIM taxa de criminalidade per capita por cidade• Proporção ZN de terreno residencial zoneada para lotes acima de 25.000 pés quadrados.• Proporção INDUS de acres comerciais não varejistas por cidade• Variável fictícia CHAS Charles River (= 1 se a área limita o rio; 0 caso contrário)• Concentração de NOX óxidos nítricos (partes por 10 milhões)• RM número médio de divisões por habitação• Proporção AGE de unidades ocupadas pelo proprietário construídas antes de 1940• Distâncias ponderadas DIS para cinco centros de emprego de Boston• Índice RAD de acessibilidade às rodovias radiais• TAX taxa de imposto de propriedade de valor total por \$ 10.000• Proporção aluno-professor PTRATIO por cidade• $B = 1000(B_k - 0,63)^2$ onde B_k é a proporção de negros por cidade• LSTAT % status mais baixo da população• MEDV Valor médio de residências ocupadas pelo proprietário em US\$ 1.000
Valores de atributos ausentes:	Nenhum
O Criador:	Harrison, D. e Rubinfeld, DL