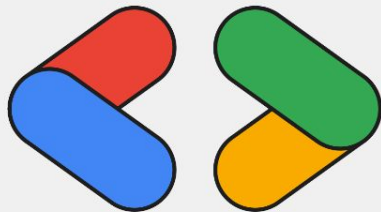


# Autópsia de IA: Hackeando e Debugando as Decisões com XAI

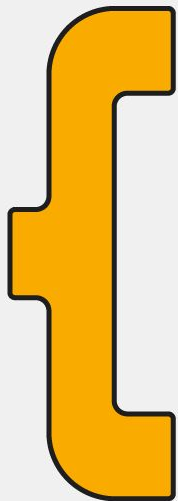
2025



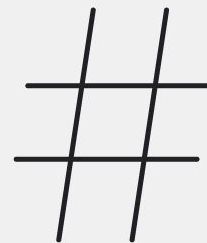
Google  
Developer  
Groups



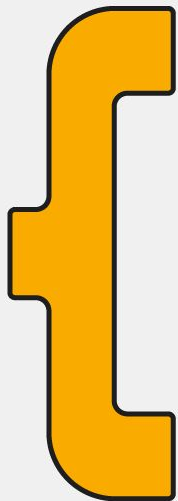
# Table of Contents



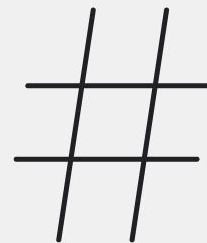
<b>Overview</b>	<b>4</b>
Quem somos nós	5
Conteúdo programático	7
Setup Check	11
Por que xAI? O problema da Caixa-Preta	14
Nosso desafio: O Titanic	15
<b>A Teoria</b>	<b>17</b>
LIME	18
SHAP	21
Quando usar cada um?	22



# Table of Contents

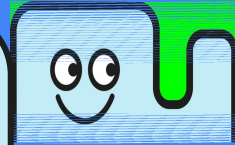


<b>Hands-On</b> .....	<b>4</b>
Setup & Treino do Modelo .....	<b>5</b>
Hackeando com LIME .....	<b>7</b>
Hackeando com SHAP .....	<b>11</b>
<b>Conclusão</b> .....	<b>17</b>
O que descobrimos? (Debriefing) .....	<b>18</b>
Q&A .....	<b>21</b>

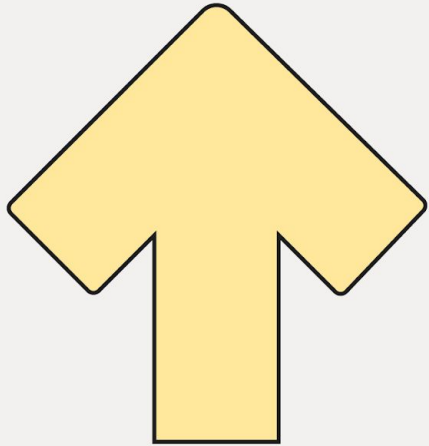


**DevFest**

Overview



01



**Quem somos  
nós?** 

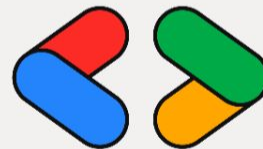


Google  
Developer  
Groups

# { DevFest }

Presidente Prudente

2025



Google  
Developer  
Groups



Bacharel em Ciência da  
Computação, Mestrando em  
Inteligência Computacional,  
Desenvolvedor Backend no  
Luizalabs, Contribuidor GDG

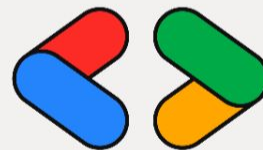
Mateus Mendonça

@mateusememe 

# { DevFest }

Presidente Prudente

2025



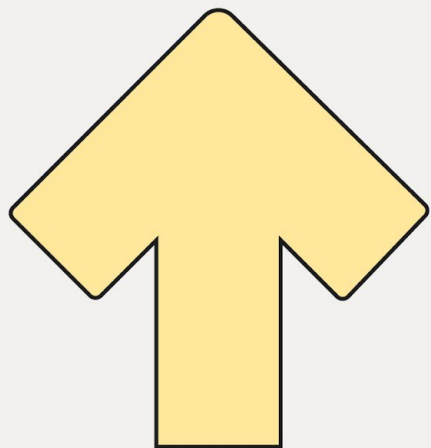
Google  
Developer  
Groups



Bacharel em Sistemas de Informação,  
Líder de Ciência de Dados na Vericode,  
Embaixadora Women Techmakers, Pós  
graduada em Engenharia de Dados e  
Data Science

Beatriz França

02



# Conteúdo Programático



Google  
Developer  
Groups



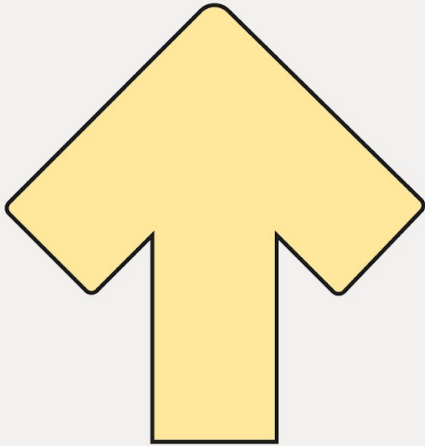
# O que vamos ver hoje ✨

- ✓ **Introdução:** O que são modelos "caixa-preta" e por que a IA Explicável (xAI) é crucial.
- ✓ **Fundamentos:** Teoria por trás das técnicas LIME e SHAP, suas vantagens e quando usá-las.
- ✓ **Prática (Hands-On):** Treinamento de um modelo com o dataset Titanic e aplicação de LIME e SHAP para gerar explicações visuais e insights.
- ✓ **Síntese:** Interpretação dos resultados, discussão sobre ética em IA e próximos passos na área de xAI.



Google  
Developer  
Groups

03



# Set-up Check



Google  
Developer  
Groups

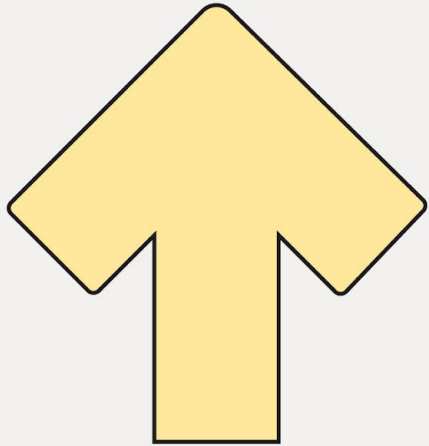
# Verificação do que vamos utilizar de ferramental

- ❑ Conta do google
- ❑ Google Colab



Google  
Developer  
Groups

04



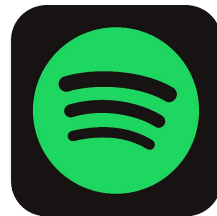
# Por que xAI? O Problema da Caixa-Preta



Google  
Developer  
Groups

# O Contrato que você não leu

Você usa IA todos os dias. Ela Recomenda suas músicas, seus filmes e até decide se o seu cartão de crédito é aprovado. Mas...você sabe o POR QUÊ ela toma essas decisões?



Google  
Developer  
Groups

# A Caixa-Preta (The Black Box)

Você usa IA todos os dias. Ela Recomenda suas músicas, seus filmes e até decide se o seu cartão de crédito é aprovado. Mas...você sabe o POR QUÊ ela toma essas decisões?



Google  
Developer  
Groups

# Quando a caixa-preta falha (e é perigoso)

**Crédito Negado:** Um algoritmo nega crédito a uma pessoa qualificada. Por quê?

**Diagnóstico Médico Errado:** Uma IA identifica um câncer de pele, mas na verdade ela só aprendeu a associar a presença de uma régua na foto com malignidade.

**Carro Autônomo:** Um carro autônomo precisa decidir entre dois cenários de colisão. Como ele tomou essa decisão ética?

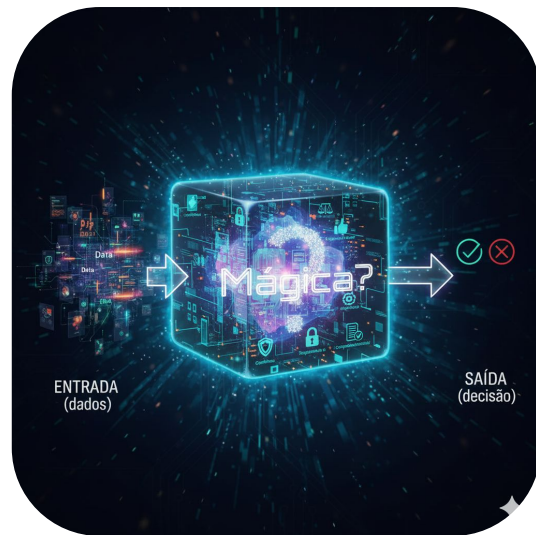


Google  
Developer  
Groups

# A Revolução da transparência

XAI não é sobre diminuir a precisão dos modelos, é sobre adicionar uma camada de interpretabilidade sobre eles.

- **Confiança**
- **Detecção de Bias (Vieses)**
- **Segurança e Robustez**
- **Conformidade Regulatória (LGPD, GDPR)**



Google  
Developer  
Groups



# Nosso desafio O TITANIC

Vamos usar um dos datasets mais famosos para treinar um modelo '**caixa-preta**' e depois usar XAI para '**hackear**' suas decisões.

**Nossa missão:** entender por que o modelo acha que uma pessoa sobreviveria ou não ao desastre do Titanic.



Google  
Developer  
Groups

# O que vamos construir

1. Treinar um modelo de alta performance (LightGBM).
2. Usar LIME para explicações locais (decisão a decisão).
3. Usar SHAP para explicações locais E globais (entender o modelo como um todo).



Google  
Developer  
Groups

**DevFest**

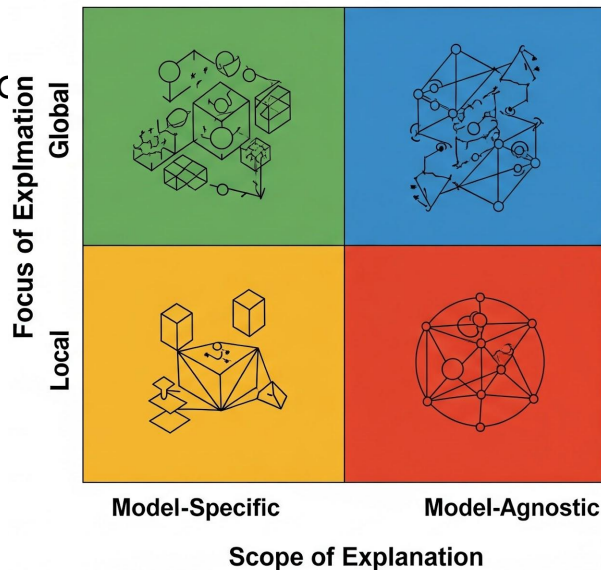
A Teoria



# A Caixa de Ferramentas do xAI

**Model-Specific vs Model-Agnostic:** vamos focar no Model-Agnóstico, que é mais versátil.

**Global vs Local:** Vamos usar *LIME* para local e *SHAP* para global.



Google  
Developer  
Groups

# LIME (Local Interpretable Model-agnostic Explanations)

❑ Imagine que você não entende como um carro de Fórmula 1 funciona (o modelo caixa-preta). Para entender como ele faz uma curva específica (uma previsão), você constrói um kart simples (modelo linear) que imita o comportamento do carro de F1 *apenas naquela curva*. O LIME faz exatamente isso com dados.

- **Passo a Passo:**

- Pega uma única previsão que você quer entender.
- Cria uma "vizinhança" de dados ligeiramente modificados ao redor desse ponto.
- Pede ao modelo caixa-preta para prever o resultado para todos esses novos pontos.
- Treina um modelo simples e interpretável (ex: regressão linear) que aprende a mapear os dados da vizinhança às previsões do modelo complexo.
- As características mais importantes nesse modelo simples são a "explicação".



Google  
Developer  
Groups

# LIME (Local Interpretable Model-agnostic Explanations)

- **Prós e Contras**

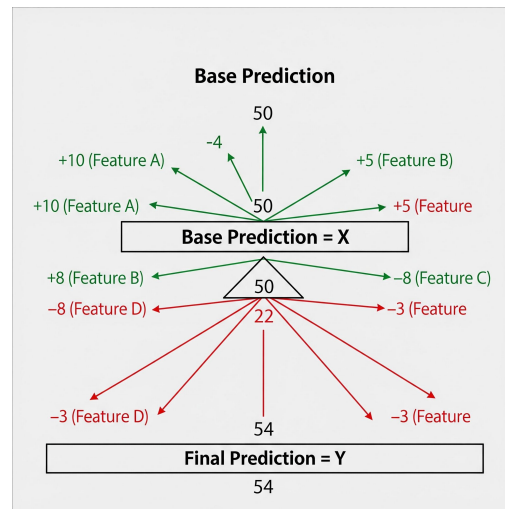
- **Vantagens:** Muito intuitivo, fácil de entender, funciona com qualquer tipo de modelo (agnóstico), explica dados tabulares, textos e imagens.
- **Desvantagem:** A definição de "vizinhança" pode ser instável, as explicações podem mudar muito com pequenas alterações, não dá uma visão geral do comportamento.



Google  
Developer  
Groups

# SHAP (SHapley Additive exPlanations)

- **Fundamentação:** Baseado na Teoria dos Jogos (Valores de Shapley), um método com forte base matemática para distribuir os "créditos" de uma previsão entre as "jogadoras" (as features).
- **Analogia:** "Imagine um time de futebol (as features) que marcou um gol (a previsão). Quanto do crédito pelo gol vai para o atacante, quanto para o meio-campista que deu o passe, e quanto para o zagueiro que iniciou a jogada? O SHAP calcula a contribuição marginal de cada 'jogador' de forma justa."
- **Conceito Central:** O valor SHAP de uma feature é a sua contribuição média para a previsão, considerando todas as combinações possíveis de outras features.



Google  
Developer  
Groups

# SHAP (SHapley Additive exPlanations)

- **Prós e Contras**

- **Vantagens:** Base matemática sólida (garante consistência e acurácia local), fornece tanto explicações locais (force plots) quanto globais (summary plots), geralmente considerado o estado da arte para dados tabulares.
- **Desvantagens:** Pode ser computacionalmente lento para grandes datasets, a interpretação dos gráficos mais complexos requer um pouco de prática.



Google  
Developer  
Groups



# LIME vs SHAP - Quando usar cada um?

- **Use LIME quando:** Você precisa de uma explicação rápida e intuitiva para uma única previsão para um público não-técnico. Ótimo para um "sanity check" rápido.
- **Use SHAP quando:** Você precisa de robustez, consistência e uma visão completa do modelo. Essencial para auditorias de modelos, detecção de bias e para entender a interação complexa entre features.



Google  
Developer  
Groups

# Preparando o Ambiente

**Chega de teoria. Hora de codar!**

Abram o Google Colab

+

Intervalo de 30min



Google  
Developer  
Groups

**DevFest**

Hands-On

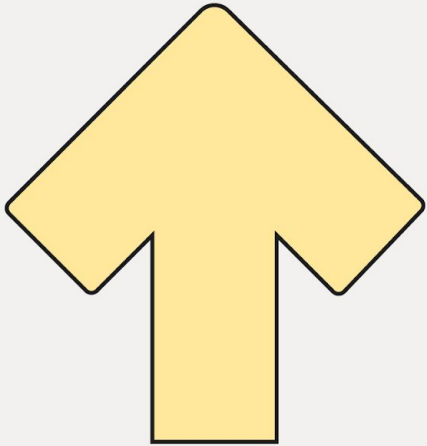


**DevFest**

Síntese e Conclusões



0X



**O que  
descobrimos?**



Google  
Developer  
Groups

# Descobertas

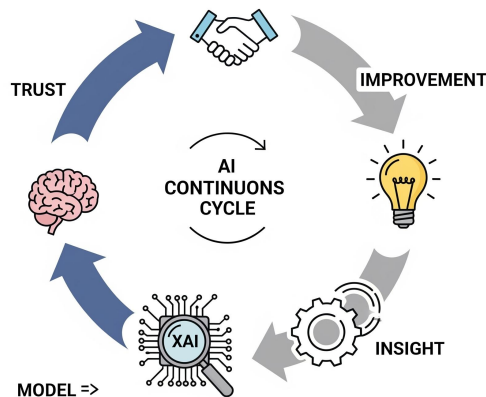
- Vimos que nosso modelo aprendeu regras historicamente conhecidas: ser mulher ('Sex\_female') ou criança ('Age') aumentava drasticamente a chance de sobrevivência.
- Descobrimos com o SHAP que a tarifa paga ('Fare') também era um fator importante, mas seu impacto não era linear.
- Com o LIME, conseguimos explicar caso a caso por que uma pessoa como 'Rose' (jovem, mulher, primeira classe) teria uma alta probabilidade de sobreviver.



Google  
Developer  
Groups

# O Poder e Importância da Explicação

- **Para Desenvolvedores:** Debugar o modelo, entender por que ele está errando, e ter mais confiança no deploy.
- **Para Negócios:** Aumentar a adoção de IA, cumprir regulações, construir confiança com o cliente.
- **Para a Sociedade:** Lutar contra vieses algorítmicos e garantir justiça e equidade.



Google  
Developer  
Groups

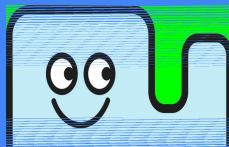
# O Futuro é transparente

- **Tendências em xAI:**
  - Explicações em tempo real.
  - Explicações em GenAI/LLM.
  - Ferramentas mais amigáveis (Low-code XAI).
  - Regulamentações de IA exigindo transparência.



Google  
Developer  
Groups





**DevFest**

**Obrigado!**

Q/A?



Mateus  
Mendonça



Maria Beatriz  
França