



SHAP

explicabilidade
de modelos para
decisões mais
inteligentes

dhauz®

SHAP: explicabilidade de modelos para decisões mais inteligentes

A inteligência artificial ocupa cada vez mais espaço nas decisões estratégicas das empresas. Modelos preditivos já influenciam desde o risco de crédito até o churn de clientes e a manutenção de máquinas. Mas, junto a esse avanço, surge o desafio de **traduzir modelos complexos para uma linguagem que gestores e executivos entendam.**

Nesse sentido, o SHAP, uma das ferramentas mais poderosas da área de Explainable AI (XAI), fica responsável por tornar inteligível aquilo que, muitas vezes, parece uma “caixa-preta”.

O ambiente de negócios mudou. A adoção de IA cresceu, os times são cada vez mais multidisciplinares e os executivos precisam tomar decisões baseadas em modelos estatísticos que eles nem sempre dominam.

Por isso, ferramentas como o SHAP se tornam fundamentais. Elas permitem:

- **Confiar no modelo:** ao entender por que uma previsão foi feita, líderes tomam decisões mais seguras.
- **Traduzir complexidade em clareza:** insights técnicos passam a ser apresentados de modo visual, direto e compreensível.
- **Direcionar ações com precisão:** em vez de campanhas amplas e genéricas, as áreas de negócio atuam de forma focada.
- **Detectar erros e viés algorítmico** antes da implantação, reduzindo riscos estratégicos, financeiros e regulatórios.

Em resumo: modelos excelentes não bastam — é preciso explicá-los.

O que é SHAP?

SHAP (SHapley Additive exPlanations) é um método de explicabilidade inspirado na **Teoria dos Jogos**, criado para explicar previsões de qualquer modelo de Machine Learning.

Ele responde a uma pergunta simples, mas poderosa:

quais fatores realmente influenciaram essa previsão?

Isso o torna parte central da área de Explainable AI (XAI), cujo objetivo é permitir que humanos entendam como algoritmos chegam às suas conclusões.

Um exemplo prático

Imagine um modelo que identifica uma transação como fraude.

Sem explicabilidade, tudo o que recebemos é o rótulo final.

Com SHAP, porém, a resposta ganha contexto:

“Classificado como fraude devido ao valor elevado, à localização incomum e ao horário da operação.”

Ao revelar por que a decisão foi tomada, o método transforma o resultado em um insight acionável para o cliente, trazendo transparência e confiança à previsão.

O que o SHAP responde dentro de uma pipeline de IA?

As métricas clássicas — Acurácia, AUC, F1 — dizem **quão bem** o modelo performa. O SHAP diz **por que** ele performa daquele jeito.

Ele entrega explicações em dois níveis:

Global

- Quais variáveis são mais importantes para o modelo como um todo?

Local

- Por que este cliente foi classificado como churn?
- O que levou esta transação a ser considerada fraude?

E funciona com qualquer tipo de modelo, embora seja especialmente eficiente para os baseados em árvores (XGBoost, LightGBM, RandomForest).

Como funciona na prática

O fluxo básico de uso do SHAP envolve quatro etapas:

- 1. Treinar o modelo** — qualquer modelo de classificação ou regressão.
- 2. Criar o explicador**
 - TreeExplainer: otimizado para modelos de árvore
 - KernelExplainer: aplicável a qualquer modelo (mais lento)
- 3. Calcular os SHAP values** — que representam a contribuição de cada variável.
- 4. Interpretar os resultados** — sempre na escala de log-odds para classificação.

A lógica é simples:

- **SHAP positivo** → empurra a previsão para a classe positiva (ex.: Fraude, Churn).
- **SHAP negativo** → empurra para a classe negativa, o oposto do objetivo de previsão.

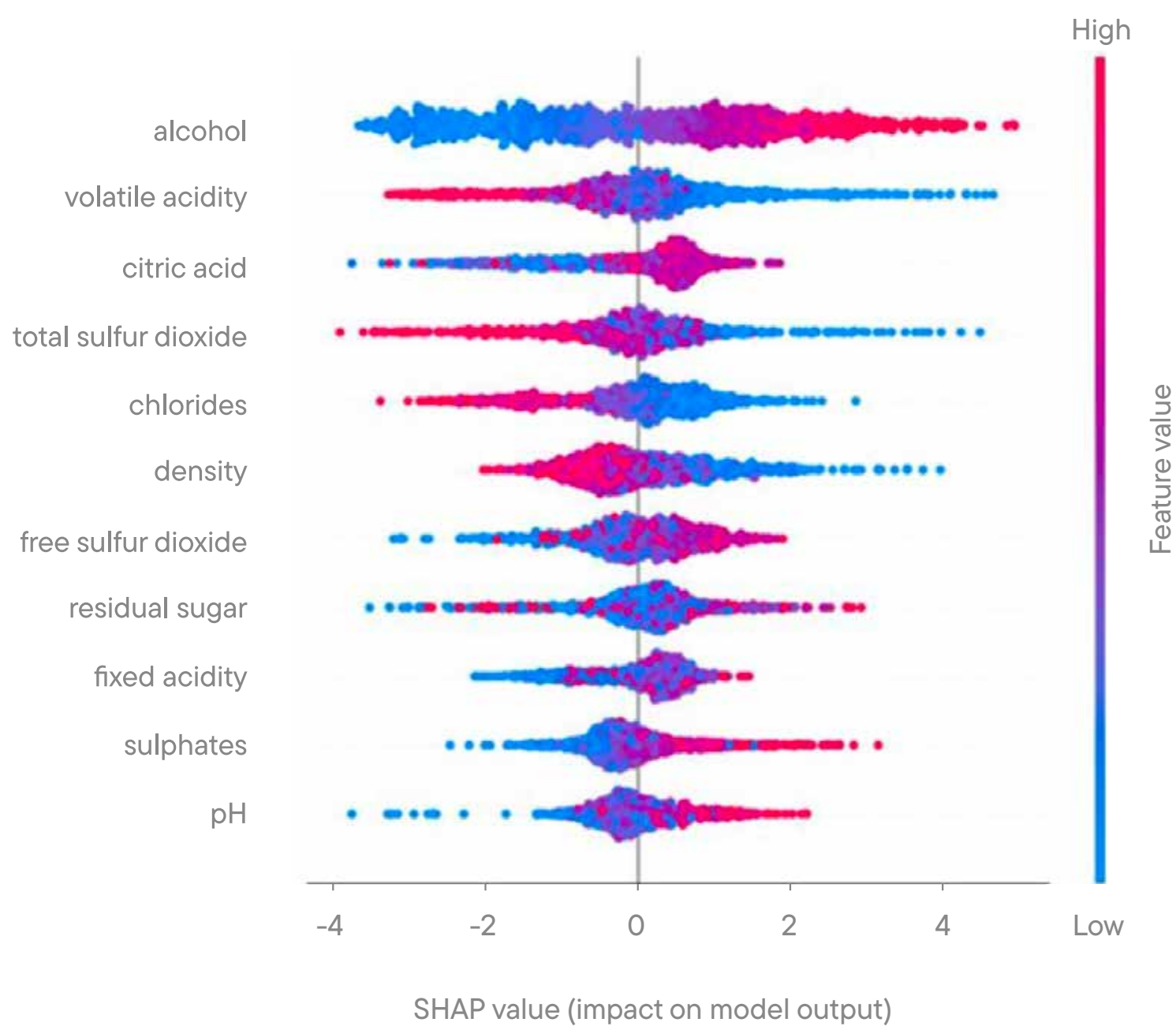
Os principais gráficos do SHAP

Summary Plot (Beeswarm)

Mostra, globalmente, as variáveis mais importantes e como elas influenciam as previsões.

É a melhor visão geral do modelo.

Previsão de qualidade do vinho vermelho



Como ler o gráfico?

Mostra as features mais importantes (globalmente) no eixo Y, ordenadas de cima pra baixo por relevância. Neste exemplo, identificamos que o alcohol é uma das features mais relevantes do modelo, enquanto o pH é menos relevante.

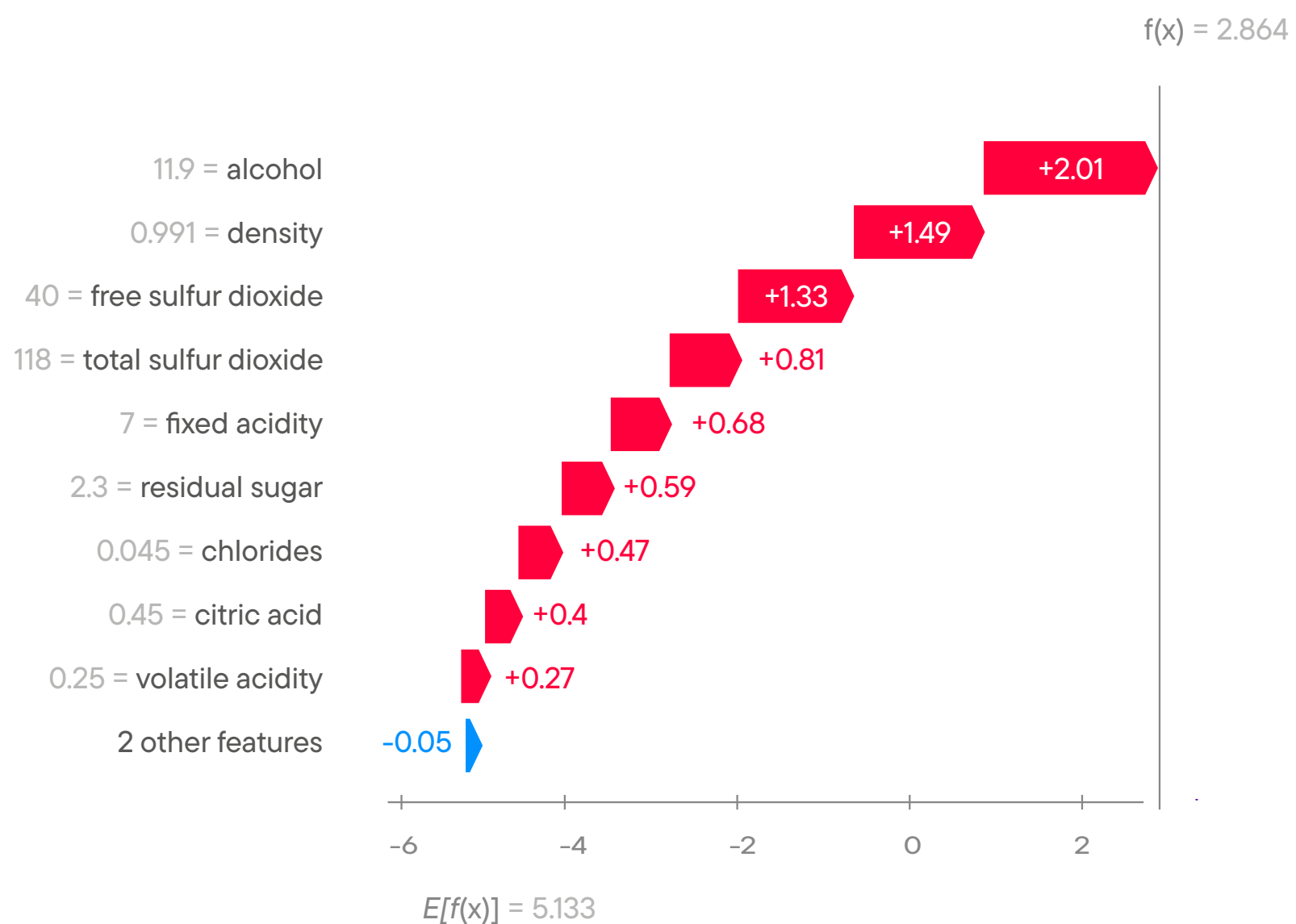
No eixo X, mostra o SHAP_values de cada ponto de dados, permitindo ver não só a importância, mas como a feature impacta. Por exemplo, pode mostrar que valores altos da variável alcohol tem SHAP_values positivos (empurram para previsão positiva), enquanto valores baixos (azul) têm SHAP_values negativos (empurram para previsão negativa).

Waterfall Plot

Explica uma única previsão.

Mostra como cada variável aumentou ou reduziu o resultado final — perfeito para esclarecer casos individuais. .

Previsão de qualidade do vinho vermelho



Como ler o gráfico?

Ele explica uma única previsão (interpretabilidade local).

Mostra como o resultado se afasta do valor base — a média prevista pelo modelo — até chegar ao valor final.

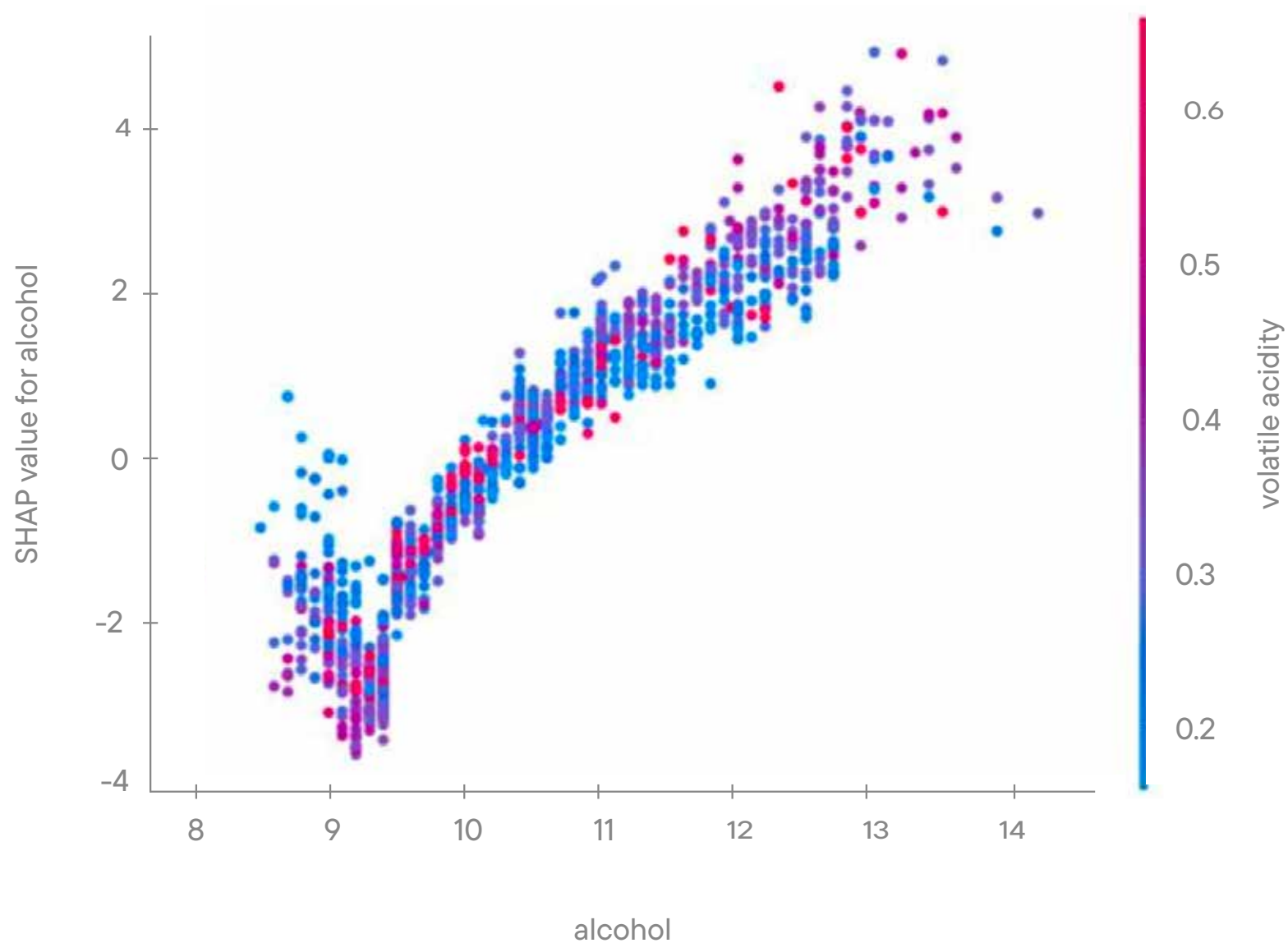
Cada feature atua como uma barra que desloca a previsão: para cima (vermelho) ou para baixo (azul), até atingir a saída final em log-odds.

Neste exemplo, o álcool e a densidade foram os fatores que mais influenciaram a previsão da qualidade do vinho tinto.

Dependence Plot

Revela relações não-lineares entre uma variável e o resultado, além de interações entre variáveis que o modelo capturou.

Previsão de qualidade do vinho vermelho



Como ler o gráfico?

Mostra o efeito de uma única feature sobre o SHAP_values em todo o dataset.

É excelente para capturar relações não-lineares e, crucialmente, efeitos de interação. O gráfico pode colorir os pontos com base em uma segunda feature, revelando se o impacto da feature A muda dependendo do valor da feature B.

Neste exemplo identificamos que a combinação de maiores valores na feature alcohol e maiores valores na feature volatile acidity aumentam a proporção de qualidade do vinho vermelho.

A base matemática por trás

O SHAP é baseado nos **Valores de Shapley**, criados por Lloyd Shapley (Prêmio Nobel, 1953).

Nesse conceito, cada variável do modelo é tratada como um “jogador” e a previsão final como o “prêmio”. O SHAP calcula quanto cada variável contribuiu para esse prêmio, considerando todas as combinações possíveis.

Ele é o único método de explicabilidade que satisfaz simultaneamente:

1. Acurácia local

A soma dos valores SHAP sempre explica exatamente a previsão.

2. Ausência (Missingness)

Se a variável não aparece, sua contribuição é zero.

3. Consistência

Se o impacto de uma variável aumenta no modelo, seu SHAP também aumenta.

Essas propriedades garantem explicações justas, estáveis e matematicamente sólidas.

Aplicações em decisões estratégicas

1. Risco de crédito

- **Sem SHAP:** “crédito negado por score baixo”.
- **Com SHAP:** a recusa se deve a “alta relação dívida/renda” e “histórico recente de atraso”.

Resultado: o cliente recebe orientação clara e a instituição reduz reclamações e retrabalho.

2. Manutenção preditiva

- **Sem SHAP:** “falha iminente na máquina 3”.
- **Com SHAP:** 90% da previsão vem do aumento da vibração do sensor A e da queda de pressão no sensor B.

Resultado: equipes focam exatamente onde está o problema.

3. Detecção e correção de viés algorítmico

SHAP e conformidade com a LGPD

A LGPD garante ao cidadão o direito de pedir explicações sobre decisões automatizadas (Art. 20).

Empresas que usam modelos preditivos precisam ser capazes de justificar suas previsões.

O SHAP permite exatamente isso.

Diante de uma contestação, a empresa pode mostrar — de forma clara e auditável — quais fatores levaram à decisão, fortalecendo transparência e confiança.

Conclusão

O SHAP não é apenas uma ferramenta técnica; é um elo essencial entre inteligência artificial e estratégia de negócios. Ele transforma modelos complexos em histórias claras, acionáveis e confiáveis, permitindo que empresas tomem decisões melhores, mais rápidas e mais transparentes.



dhauz.®

