



MINERAÇÃO DE DADOS

Prof. Me. Napoleão Póvoa Ribeiro Filho



GOVERNO DO
TOCANTINS
TRABALHANDO E CUIDANDO DE TODOS

MÉTRICAS DE AVALIAÇÃO DE MODELOS DE MACHINE LEARNING

Entender como medir a qualidade de um modelo além da acurácia

POR QUE AVALIAR MODELOS

- Avaliar é essencial para saber se o modelo é útil
- Métrica errada pode levar a decisões ruins
- Cada problema exige uma métrica adequada
- Por exemplo, acurácia alta pode enganar em problemas com classes desbalanceadas

MATRIZ DE CONFUSÃO

- Visualiza acertos e erros por classe
- Base para calcular todas as outras métricas

	Predito: Positivo	Predito: Negativo
Real: Positivo	VP	FN
Real: Negativo	FP	VN

ACURÁCIA

- Mede a proporção total de acertos do modelo
- Boa em classes balanceadas
- Pode ser enganosa se uma classe domina os dados

$$\text{Acurácia} = \frac{\text{Nº de acertos}}{\text{Total de amostras}}$$

PRECISÃO (PRECISION)

- Mede quantas previsões positivas foram realmente corretas
- "Entre todos os que o modelo disse que eram positivos, quantos realmente eram?"
- Útil quando o custo de falsos positivos é alto (ex: spam, fraude)

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Positivos}}$$

EXEMPLOS

Filtro de spam em e-mails corporativos

- Se o sistema marcar um e-mail importante (como uma proposta de contrato) como spam, o usuário pode perder uma oportunidade de negócio.
- Esse é um falso positivo, e o custo é altíssimo.
- Já deixar passar um spam ocasional (falso negativo) é menos grave — o usuário pode simplesmente apagá-lo.

REVOCAÇÃO (RECALL)

- Mede quantos casos positivos reais foram identificados
- "Entre todos os que realmente eram positivos, quantos o modelo conseguiu encontrar?"
- Importante quando não identificar um positivo é crítico

$$\text{Revocação} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Negativos}}$$

EXEMPLOS

Diagnóstico inicial de uma doença grave (câncer, HIV, Covid-19)

- Se o sistema errar e disser que o paciente não tem a doença quando, na verdade, ele tem (falso negativo), o tratamento pode ser adiado ou nem iniciado, com risco de morte.
- Já um falso positivo (dizer que tem a doença quando não tem) leva apenas a exames adicionais para confirmar o diagnóstico.

EXEMPLOS

Diagnóstico inicial de uma doença grave (câncer, HIV, Covid-19)

- Se o sistema errar e disser que o paciente não tem a doença quando, na verdade, ele tem (falso negativo), o tratamento pode ser adiado ou nem iniciado, com risco de morte.
- Já um falso positivo (dizer que tem a doença quando não tem) leva apenas a exames adicionais para confirmar o diagnóstico.

O QUE SERIA NESSE CASO?

Deteção de discurso de ódio em redes sociais (discurso de ódio ou inofensivas)

- Muitos falsos positivos → muitas mensagens de ódio passam
- Muitos falsos negativos → muitas mensagens inofensivas serão marcadas como discurso de ódio

F1 SCORE

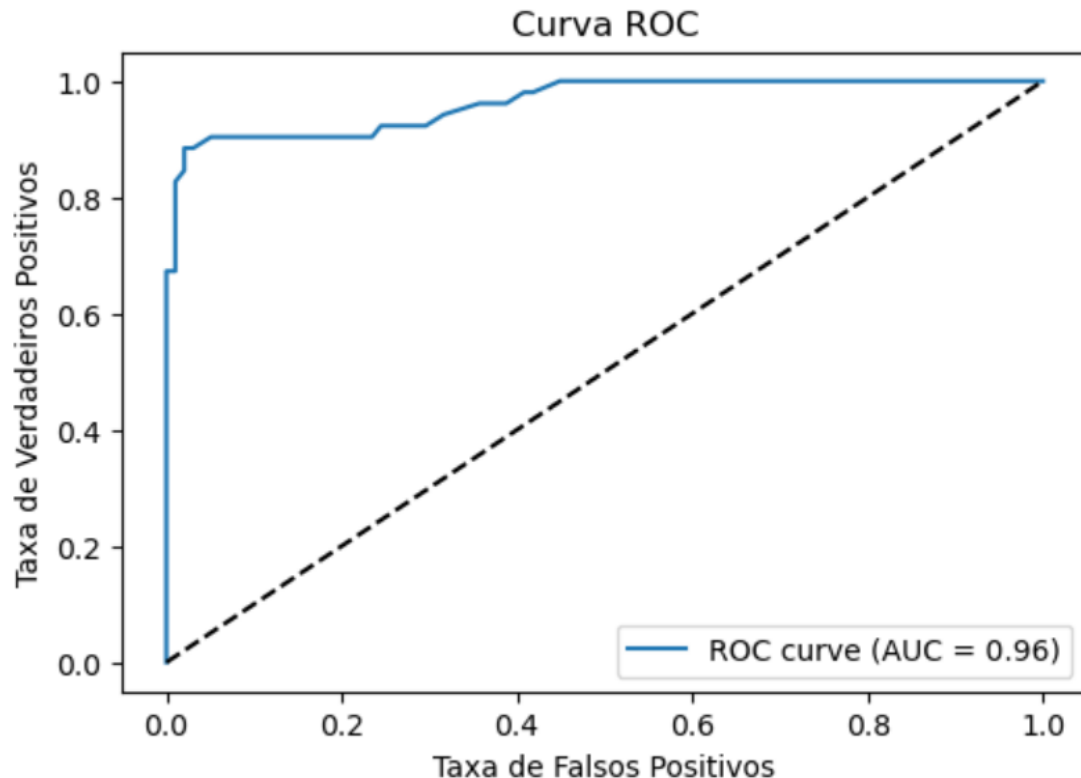
- Combina Precisão e Revocação em uma única métrica
- Equilíbrio entre detectar e prever corretamente
- Útil em dados desbalanceados
- Só será alto se ambos forem altos

$$F1 = 2 \cdot \frac{\text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}}$$

CURVA ROC (RECEIVER OPERATING CHARACTERISTIC)

- Um gráfico que representa o desempenho do classificador em diferentes limiares de decisão
 - Eixo X → Taxa de Falsos Positivos (FPR) = $FP / (FP + TN)$
 - Eixo Y → Taxa de Verdadeiros Positivos (TPR ou Recall) = $TP / (TP + FN)$
- A métrica associada a ela é a AUC (Area Under the Curve) — Área sob a Curva ROC.

CURVA ROC (RECEIVER OPERATING CHARACTERISTIC)



- Linha tracejada Representa um modelo aleatório, com $AUC = 0.5$
- Tudo acima da linha mostra que o modelo é melhor que o acaso.
- $AUC = 0.96$ significa que o modelo tem um desempenho excelente em distinguir as classes.
- Valor do AUC:
 - 1 \rightarrow classificação perfeita
 - 0.7 a 0.8 \rightarrow desempenho aceitável
 - $5 > AUC < 6 \rightarrow$ baixa capacidade discriminatória
 - 0.5 \rightarrow desempenho aleatório
 - $< 0.5 \rightarrow$ erra mais do que acerta

CURVA PRECISÃO vs REVOCAÇÃO (RECALL)

- Ferramenta fundamental para avaliar classificadores, especialmente em problemas com classes desbalanceadas
- Mostra o equilíbrio entre precisão e revocação conforme o limiar de decisão muda
- Quando uma sobe, a outra geralmente cai
- Ajuda a escolher o melhor ponto de corte (threshold)

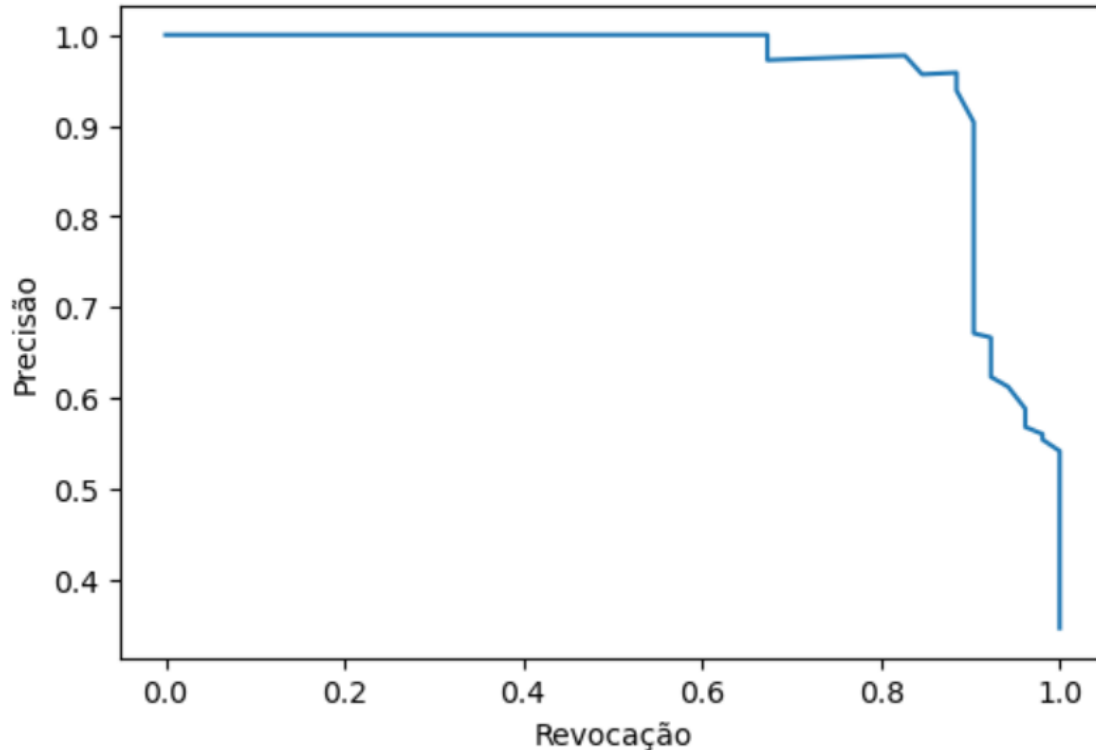
EXEMPLO

Imagine um modelo que detecta fraudes bancárias

- Se aumentar a sensibilidade (recall), ele detecta quase todas as fraudes, mas também marca muitos clientes honestos (baixa precisão).
- Se for muito rigoroso (alta precisão), detecta apenas as fraudes mais óbvias e deixa passar várias (baixo recall).
- A curva Precisão vs Revocação mostra graficamente esse equilíbrio, ajudando a escolher o ponto ideal entre não deixar passar fraudes e não gerar falsos alarmes demais.

CURVA PRECISÃO vs REVOCAÇÃO

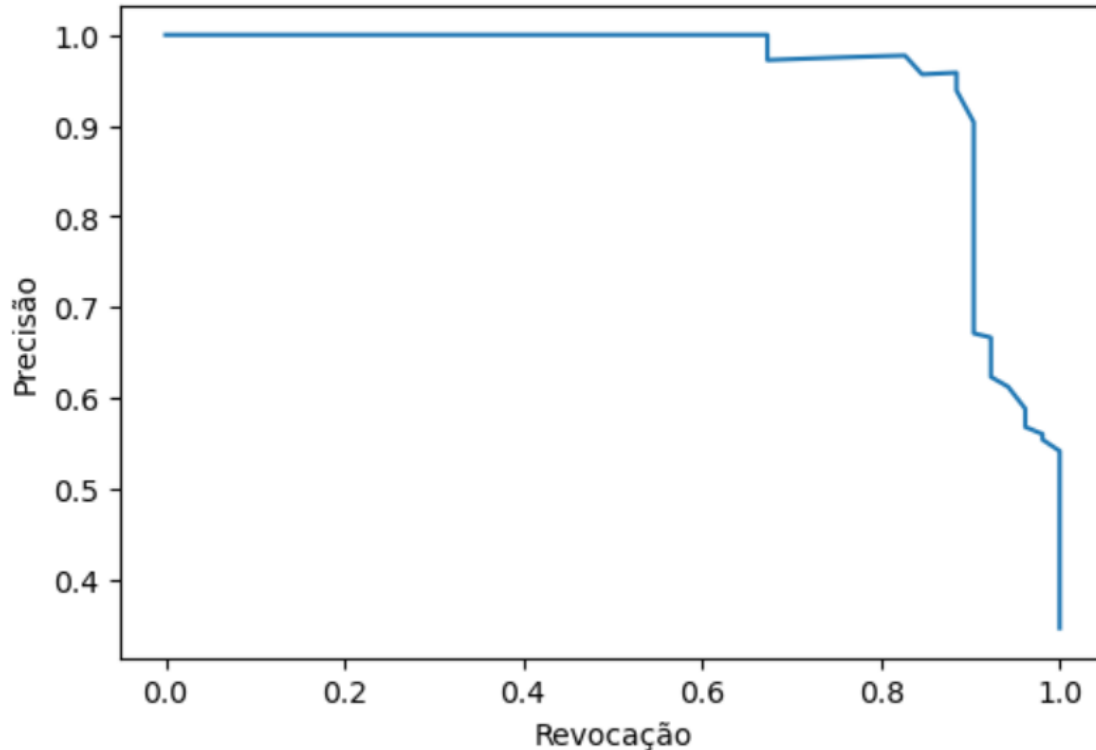
Curva Precisão vs Revocação



- A curva tem alta precisão (>0.9) na maior parte do gráfico. Isso indica que o modelo quase não erra quando prevê positivo.
- Só quando a revocação se aproxima de 1 (detectando quase todos os positivos) é que a precisão cai abruptamente.
- Esse comportamento é típico de um modelo robusto, que mantém equilíbrio entre precisão e recall até o limite.

CURVA PRECISÃO vs REVOCAÇÃO

Curva Precisão vs Revocação



- O modelo possui excelente desempenho, com um bom trade-off entre precisão e recall.
- É especialmente adequado para tarefas onde ambos os erros (falsos positivos e falsos negativos) importam
- O ponto ótimo costuma estar onde a curva começa a cair rapidamente, pois ali temos alto recall sem comprometer muito a precisão.

RESUMO

REGIÃO DA CURVA	INTERPRETAÇÃO
Parte inicial (precisão alta, recall baixo)	O modelo é conservador e só acerta os casos mais óbvios
Região intermediária	Melhor equilíbrio entre precisão e recall
Final (recall alto, precisão baixa)	O modelo tenta detectar tudo, mas com muitos falsos positivos

COMPARATIVO ENTRE AS MÉTRICAS

Métrica	Ideal para
Acurácia	Classes balanceadas
Precisão	Evitar falsos positivos
Revocação	Evitar falsos negativos
F1-score	Balancear previsão e revocação
Matriz de confusão	Diagnóstico completo de erros
Curva ROC	Avaliação geral do modelo
Curva P-R	Avaliação detalhada em classes desbalanceadas

CONCLUSÃO

- Não existe métrica universal: depende do problema
- Avaliar além da acurácia é essencial
- Combine métricas para melhor entendimento
- A métrica certa leva à decisão certa.



UNITINS

UNIVERSIDADE ESTADUAL DO TOCANTINS