



Atividade Prática – Métricas de Avaliação de Modelos de Classificação

1. Considere um problema de classificação binária onde estamos tentando prever se um paciente possui uma determinada doença com base em múltiplos sintomas. As matrizes a seguir apresentam os resultados de classificação de dois modelos.

Modelo 1:

Positivo	150	100
Negativo	20	550

Verdadeiro Positivo (VP) = 150
Falso Positivo (FP) = 100
Falso Negativo (FN) = 20
Verdadeiro Negativo (VN) = 550

Modelo 2:

Positivo	100	150
Negativo	70	500

Verdadeiro Positivo (VP) = 100
Falso Positivo (FP) = 150
Falso Negativo (FN) = 70
Verdadeiro Negativo (VN) = 500

- a) Considerando os modelos 1 e 2, calcule a acurácia, precisão, revocação e F1-score. Depois, analise e defina que modelo de classificação teve o melhor resultado, levando em conta o falso negativo.

Fórmulas das Métricas

- Acurácia: $(VP + VN)/(VP + FP + FN + VN)$
- Precisão: $VP/(VP + FP)$
- Revocação (Recall): $VP/(VP + FN)$
- F1-Score: $2 \cdot (Precisão \cdot Revocação)/(Precisão + Revocação)$

Cálculos do Modelo 1

- Acurácia: $(150 + 550)/(150 + 100 + 20 + 550) = 700/820 \approx 0.8549$
- Precisão: $150/(150 + 100) = 150/250 = 0.6$
- Revocação: $150/(150 + 20) = 150/170 \approx 0.8824$
- F1-Score: $2 \cdot (0.6 \cdot 0.8824)/(0.6 + 0.8824) \approx 0.7119$

Cálculos do Modelo 2

- Acurácia: $(100 + 500)/(100 + 150 + 70 + 500) = 600/820 \approx 0.7317$
- Precisão: $100/(100 + 150) = 100/250 = 0.4$
- Revocação: $100/(100 + 70) = 100/170 \approx 0.5882$
- F1-Score: $2 \cdot (0.4 \cdot 0.5882)/(0.4 + 0.5882) \approx 0.4762$

O Modelo 1 tem menos falsos negativos (FN = 20) em comparação ao Modelo 2 (FN = 70), o que é crucial no contexto médico onde falhar em identificar uma doença (falso negativo) pode ter consequências sérias.

2. Compare as vantagens e desvantagens de utilizar o F_β score em um cenário de classificação de desequilíbrio de classes. Em quais situações cada uma dessas métricas seria mais apropriada?

F_β Score: É útil quando precisão e recall têm diferentes níveis de importância. Quando falsos negativos são mais críticos (ex: diagnósticos médicos), o recall é priorizada, e um valor maior de β é usado para enfatizar o recall. E quando falsos positivos são mais críticos, a precisão é priorizada, e um valor menor de β é usado.

3. Qual a importância do ajuste de β (F_β score) em um cenário onde precisão e recall têm diferentes níveis de importância.

O ajuste de β no F_β score permite que o modelo de classificação seja avaliado de acordo com a importância relativa da precisão e do recall para o problema em questão. Este ajuste é essencial para alinhar a avaliação do modelo com os objetivos específicos do negócio ou domínio de aplicação, garantindo que o modelo atenda aos requisitos críticos de desempenho e impacto operacional.

4. O índice Kappa é frequentemente utilizada para medir a concordância entre dois classificadores. Quais são as limitações dessa métrica?

Pode ser afetado pela prevalência das classes, pode não refletir a significância prática e em casos de alto desequilíbrio de classes, Kappa pode apresentar valores baixos apesar de alta acurácia.