

## Métricas para Regressão:

### 1. MSE = Mean Squared Error (Erro Quadrático Médio):

- Mede o erro médio ao quadrado entre os valores previstos e reais.
- Penaliza fortemente erros grandes.
- Muito usado em regressão linear.

### 2. MAE = Mean Absolute Error (Erro Absoluto Médio)

- Erro médio em valor absoluto.
- Cada erro pesa igualmente.

### 3. RMSE Root Mean Squared Error (Raiz do Erro Quadrático Médio)

- É a raiz quadrada do MSE.
- Traz o erro de volta para a mesma escala do target.
- Sensível a grandes erros, mas mais interpretável que o MSE.

### 4. $R^2$ Coeficiente de Determinação

- Indica **quanto da variação do target o modelo consegue explicar**.
- Varia de 0 a 1 (ideal próximo de 1).

### 5. $R^2$ Ajustado (Adj. $R^2$ )

- Variação do  $R^2$  que **penaliza modelos com muitas variáveis**.
- Evita falsa impressão de melhora quando se adicionam features irrelevantes.
- Melhor para regressão múltipla do que o  $R^2$  simples.

## Métricas para Classificação:

### 1. Acurácia

- Percentual de previsões corretas.
- Útil quando as classes estão **balanceadas**. Não é boa para problemas desbalanceados.

### 2. Precisão (Precision)

- Entre as previsões positivas, **quantas realmente eram positivas**.
- Boa quando o custo de **falso positivo** é alto.

### 3. Recall (Sensibilidade)

- Entre os positivos reais, quantos o modelo identificou corretamente.
- Boa quando o custo de **falso negativo** é alto

### 4. F1-Score

- Média harmônica entre **precisão e recall**.
- Indicado quando o dataset é desbalanceado é necessário equilibrar FP e FN

### 5. AUC-ROC

- AUC = Área sob a curva ROC
- ROC compara **TPR vs. FPR**
- Mede a capacidade do modelo em separar as classes.
- Valores:
- 0.5 = modelo aleatório e 1.0 = perfeito
- Indica qualidade da ordenação das probabilidades.

### 6. PPS Predictive Power Score

- O PPS (Predictive Power Score) é uma métrica moderna usada para medir o poder preditivo entre duas variáveis, podendo ser usada tanto em classificação quanto em regressão.
- Ele é especialmente útil como alternativa à correlação de Pearson, principalmente quando:
  - a relação entre as variáveis não é linear
  - há variáveis categóricas
  - há mistura de tipos (numérico × categórico)
  - você quer saber se A prevê B, não apenas se estão relacionadas