

## Definições

**Acurácia:** É a razão entre as predições corretas pelo total.

$$\frac{TP + TN}{TP + TN + FP + FN}$$

**Precision:** É o número de verdadeiros positivos, dividido pelo número de positivos estimados pelo modelo.

$$\frac{TP}{TP + FP}$$

**Recall:** Dado que o estado verdadeiro é positivo, qual a proporção de verdadeiro positivo.

$$\frac{TP}{TP + FN}$$

**F1 Score:** É uma média harmônica entre o *Precision* e o *Recall*

$$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN}$$

## Quando utilizar

**Acurácia** em geral é a uma boa forma de se verificar a qualidade de um modelo, pois é uma métrica que avalia o corpo geral do resultado do seu código, porém ela considera os erros do tipo 1 e 2 com o mesmo peso, caso os tipos de erro tenham alguma importância diferente, sendo um mais impactante que o outro, não é o mais viável e recomendável.

**Precision** é uma boa opção quando o erro do tipo I é mais importante que o erro do tipo II. Define a quantidade que o modelo disse que são positivos, quantos realmente são positivos. Ele dá uma sensação de impacto.

**Recall** é uma boa opção quando o erro do tipo II é mais importante que o erro do tipo I. Define a quantidade que são positivos, quantos o meu modelo acertou. Ele dá uma sensação de captura.

**F1 Score** é uma boa forma para se verificar os erros em geral do modelo, vendo que ele faz uma média dos dois resultados anteriores. Quando os erros são mais importantes que os acertos na previsão, é uma métrica recomendável. \*OBS: Complementando com essa métrica, temos o **Fβ Score**, que simplesmente é adiciona um peso para o *Precision* no cálculo, gerando assim uma

diferença de importância entre os erros do tipo I e II.

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$

## O Recomendável Para Classes Desbalanceadas

A acurácia para dados desbalanceados não é provavelmente melhor opção, visto que nosso modelo está enviesado pela diferença de classes, ele em sua maioria vai acertar grande parte dos casos, mesmo que os erros de tipo I e II fiquem mais altos pelo padrão a ser seguido, o seu resultado será alto.

Para um modelo desbalanceado o recomendável é se observar o *F1 Score*. Já que nosso modelo está inicialmente enviesado para a classe de maior quantidade, a taxa de erros provavelmente será muito alta, fazendo assim com que nosso modelo talvez fique com essa métrica baixíssima.

Quando aplicamos um método para igualar as classes, na acurácia não iremos, provavelmente, perceber uma mudança muito nítida, mas ao compararmos o *F1 Score*, iremos verificar, provavelmente, uma mudança para melhor, visto que agora temos um modelo honesto e sem viés, assim provavelmente cometendo menos que anteriormente.