

MÉTRICAS DE AVALIAÇÃO

ACURÁCIA, PRECISÃO, *RECALL*: COMO UTILIZAR?

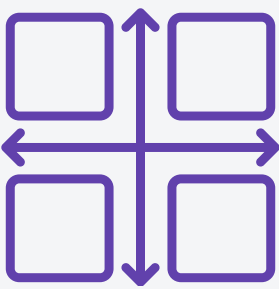


Avaliação de modelos de classificação

Ao construir um modelo de classificação, é importante avaliar seu desempenho para garantir que ele esteja funcionando conforme o esperado.

Existem várias métricas disponíveis para avaliar modelos de classificação, mas nem todas são adequadas para todas as situações.

Matriz de confusão



Uma matriz de confusão é uma tabela que mostra a distribuição de previsões corretas e incorretas de um modelo de classificação.

Ela é dividida em quatro quadrantes, cada um representando um tipo de resultado:

		Detectada	
		Sim	Não
Real	Sim	Verdadeiro positivo (VP)	Falso negativo (FN)
	Não	Falso positivo (FP)	Verdadeiro negativo (VN)

Verdadeiros positivos: classificação correta da classe positivo.

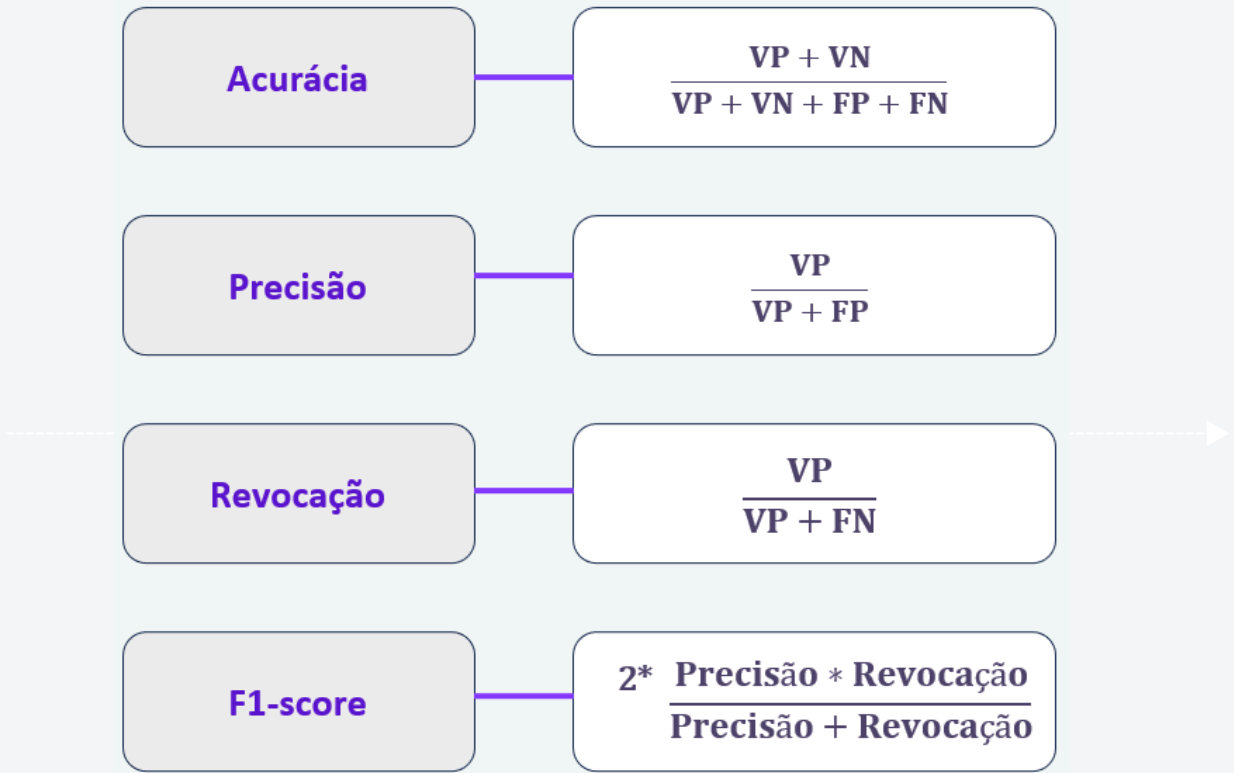
Falsos negativos (erro tipo II): erro em que o modelo previu a classe negativo quando o valor real era classe positivo.

Falsos positivos (erro tipo I): erro em que o modelo previu a classe positivo quando o valor real era classe negativo.

Verdadeiros negativos: classificação correta da classe negativo.

Métricas de avaliação

Após contar os valores da matriz de confusão, é possível calcular métricas de avaliação para a classificação.



Como aplicar cada uma delas

Acurácia

É uma métrica de avaliação de classificadores que mede a porcentagem de exemplos que foram classificados corretamente. Ela é uma boa indicação geral do desempenho de um modelo, mas pode ser enganadora em problemas de classificação desequilibrados, nos quais há muito mais instâncias de uma classe do que de outra.



- **Em um problema de detecção de fraude em cartões de crédito, onde a classe negativa (transação legal) é muito mais comum do que a classe positiva (transação fraudulenta), um modelo que simplesmente classifica todas as transações como legais teria uma acurácia muito alta, mesmo que não detectasse nenhuma fraude.**

Precisão

É uma métrica de avaliação de classificadores que mede a porcentagem de exemplos classificados como positivos que realmente são positivos. Ela é importante em problemas de classificação em que **os erros falsos positivos são mais prejudiciais do que os erros falsos negativos.**



- **Em um sistema de triagem de crédito, é importante minimizar o número de pessoas que são aprovadas para crédito, mas não são capazes de pagá-lo. Nesse caso, um modelo com alta precisão é importante para evitar a aprovação de pessoas que não seriam capazes de pagar o empréstimo.**

Revocação

É uma métrica de avaliação de classificadores que mede a porcentagem de exemplos positivos que foram classificados corretamente como positivos. Ela é importante em problemas de classificação em que **os erros falsos negativos são mais prejudiciais do que os erros falsos positivos.**



- **Em um sistema de detecção de câncer, é importante minimizar o número de cânceres que não são detectados. Nesse caso, um modelo com alta revocação é importante para garantir que todos os casos de câncer sejam identificados.**

F1-score

É uma métrica de avaliação de classificadores que combina a precisão e a revocação em uma única medida. Ele é uma média harmônica das duas métricas, o que significa que é mais influenciado pelos valores menores.



- **Se um modelo tem uma precisão de 90% e uma revocação de 80%, o seu F1-score será de 86,6%. Isso indica que o modelo tem um desempenho geral bom, pois é preciso e tem alta revocação.**