



**INSTITUTO
FEDERAL**
Norte de Minas Gerais

Introdução a Sistemas Inteligentes

Como avaliar Modelos?
Métricas para Modelos

Prof^a. Suzana Mota








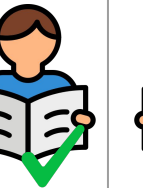
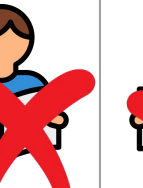













Como definir se um modelo é bom?

$$\textit{Acurácia} = \frac{\textit{Previsões Corretas}}{\textit{Total de Previsões}}$$

```
from sklearn.metrics import accuracy_score  
  
accuracy_score(y_true, y_pred)
```

LABELS

									
1	2	3	4	5	6	7	8	9	10
									

Predição do Modelo

Como definir se um modelo é bom?

$$\textit{Acurácia} = \frac{\textit{Previsões Corretas}}{\textit{Total de Previsões}}$$

$$\textit{Acurácia} = 8 / 10 = 80\%$$

Esse modelo é bom o suficiente?

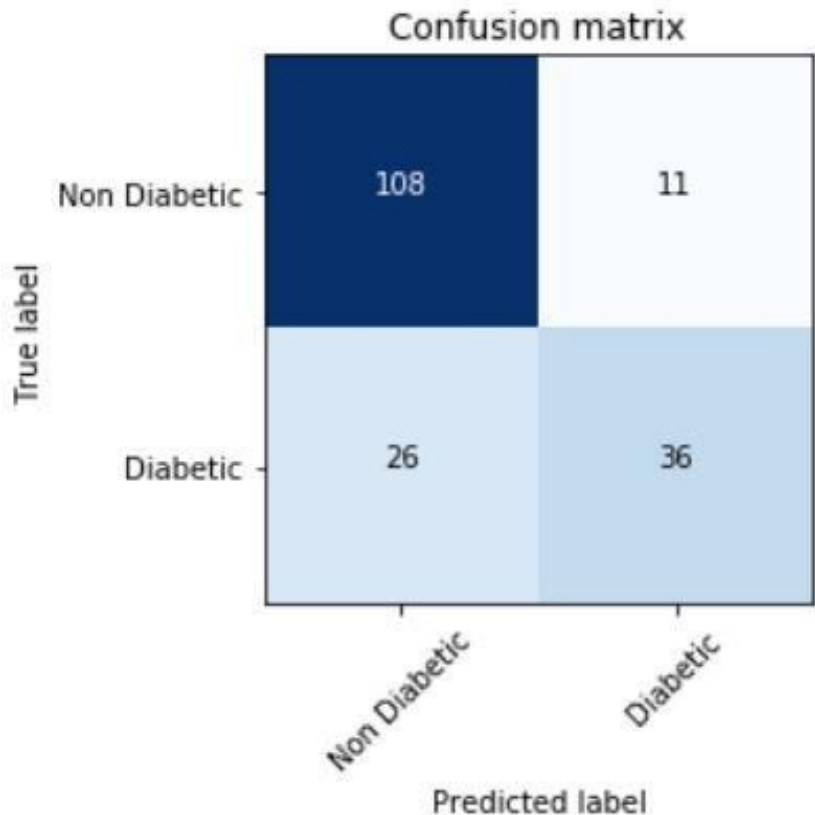
Acurácia

- É o número de acertos (positivos) dividido pelo número total de exemplos.
- **Em problemas com classes desproporcionais, ela causa uma falsa impressão de bom desempenho!**
- Por exemplo, num dataset em que 80% dos exemplos pertençam a uma classe, só de classificar todos os exemplos naquela classe como positivos já se atinge uma precisão de 80%, mesmo que todos os exemplos da outra classe estejam classificados incorretamente.

Acurácia

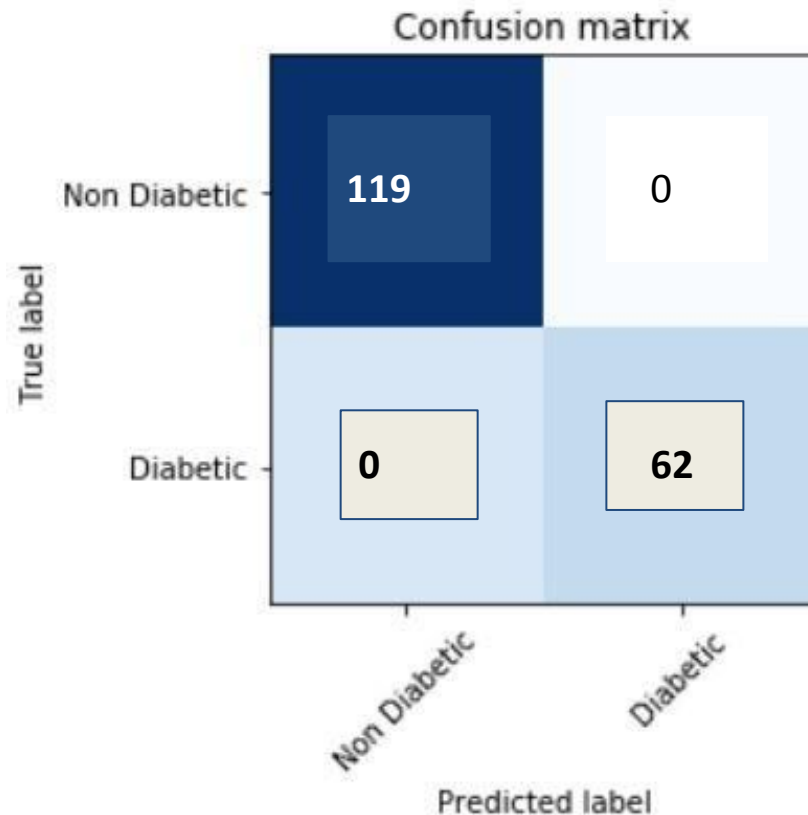
- Imagina que eu tenho um dataset com:
 - 98% de NÃO FRAUDES
 - 2% de FRAUDES
- E treinei um modelo que tem acurácia de 98%.
- Portanto ele só tem 2% de erros, mas existe o risco do meu modelo estar errando em detectar 100% das fraudes!
- Como posso saber se isso está acontecendo?

Matriz de Confusão

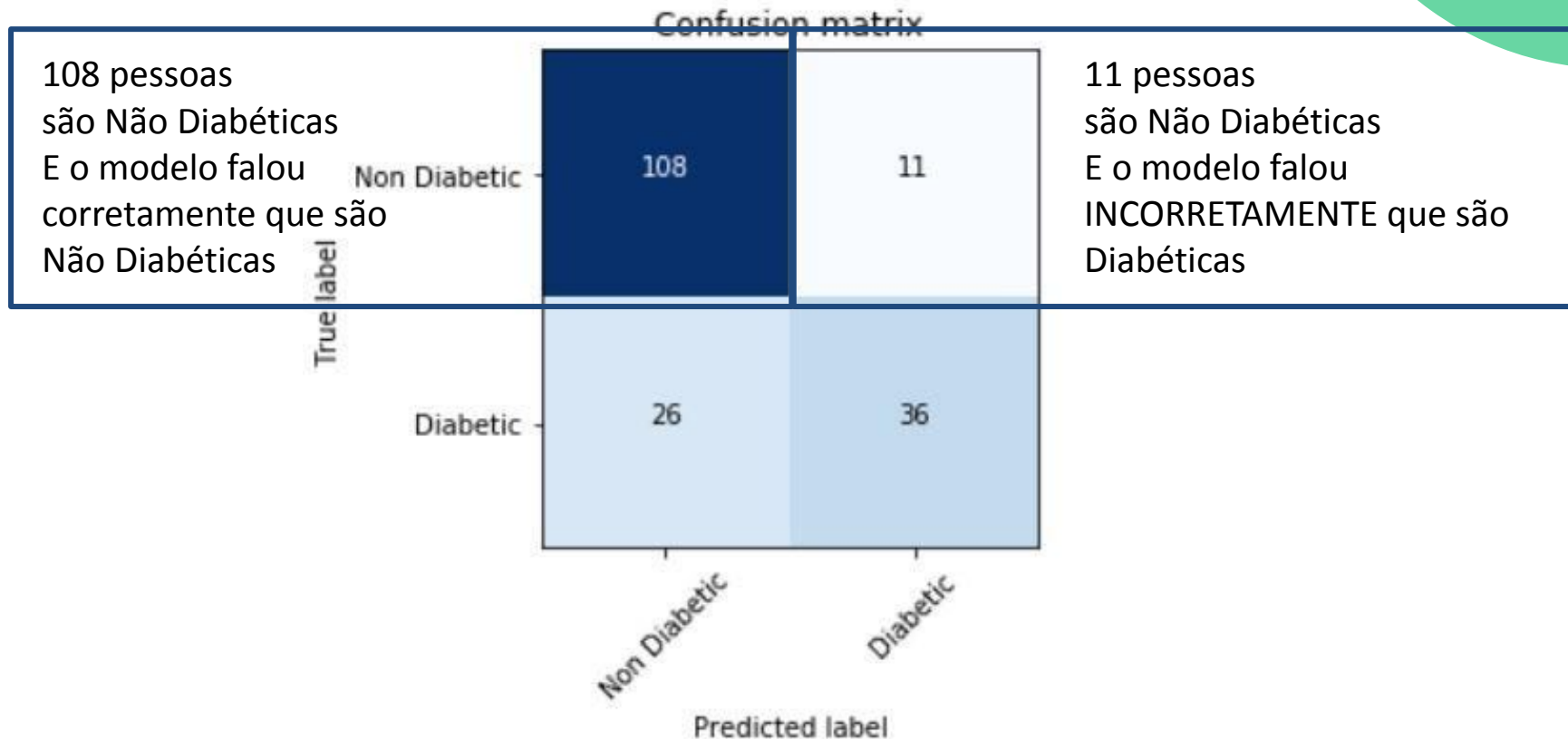


A matriz de confusão fornece uma representação tabular que resume o desempenho do modelo em termos de suas previsões em comparação com os valores reais.

Matriz de Confusão



Matriz de Confusão



Matriz de Confusão

Confusion matrix			
True label	Non Diabetic	108 pessoas são Não Diabéticas E o modelo falou corretamente que são Não Diabéticas	11 pessoas são Não Diabéticas E o modelo falou INCORRETAMENTE que são Diabéticas
	Diabetic	26 pessoas são Diabéticas E o modelo falou INCORRETAMENTE que são Não Diabéticas	36 pessoas são Diabéticas E o modelo falou CORRETAMENTE que são Diabéticas
		Non Diabetic	Diabetic
		Predicted label	

Case Fraudes

Tenho um dataset com 1.000 valores
E quero detectar a presença de fraudes

Respostas Certas

	Fraude (Positivo)	Não Fraude (Negativo)
Previsão Fraude	15	10
Previsão Não Fraude	5	970

Respostas do Modelo

Case Fraudes

	Fraude (Positivo)	Não Fraude (Negativo)
Previsão Fraude	15	10
Previsão Não Fraude	5	970

Itens em Verde:

Verdadeiro Positivos: 15 fraudes corretamente identificadas como fraudes.

Verdadeiro Negativos: 970 Transações que NÃO eram fraudes e foram identificadas corretamente como NÃO fraude.

Case Fraudes

	Fraude (Positivo)	Não Fraude (Negativo)
Previsão Fraude	15	10
Previsão Não Fraude	5	970

Itens em Vermelho:

Falsos Positivos: 10 transações que NÃO eram fraudes mas foram identificadas INCORRETAMENTE como fraude.

Falsos Negativos: 5 itens que eram fraudes e foram classificadas INCORRETAMENTE como fraude.

:

Acurácia

A acurácia mede a proporção de previsões corretas em relação ao total de previsões.

:

	Fraude (Positivo)	Não Fraude (Negativo)
Previsão Fraude	15	10
Previsão Não Fraude	5	970

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad \text{Acurácia} = \frac{15 + 970}{1000} = \frac{985}{1000} = 0,985 \text{ (ou 98,5\%)}$$

O modelo parece muito bom! Mas comete erros e deixa passar fraudes!

Precisão (Precision)

A precisão mede a proporção de verdadeiros positivos em relação ao total de positivos previstos.

:

	Fraude (Positivo)	Não Fraude (Negativo)
Previsão Fraude	15	10
Previsão Não Fraude	5	970

$$\text{Precisão} = \frac{VP}{VP + FP}$$

$$\text{Precisão} = \frac{15}{15 + 10} = \frac{15}{25} = 0,6 \text{ (ou 60\%)}$$

Dentre as fraudes que ele achou, quais eram realmente fraudes?

De todas as transações que o modelo previu como fraudulentas, apenas 60% eram realmente fraudes.

Recall (Revocação)

O recall mede a proporção de verdadeiros positivos em relação ao total de positivos reais.

	Fraude (Positivo)	Não Fraude (Negativo)
Previsão Fraude	15	10
Previsão Não Fraude	5	970

$$\text{Recall} = \frac{VP}{VP + FN}$$

$$\text{Recall} = \frac{15}{15 + 5} = \frac{15}{20} = 0,75 \text{ (ou 75\%)}$$

Dentre as fraudes que existem, quantas fraudes foram detectadas de fato?
O modelo conseguiu identificar 75% das fraudes que realmente ocorreram.

F1 Score

O F1 Score é a média harmônica entre precisão e recall. É uma métrica útil quando você deseja um equilíbrio entre precisão e recall.

$$\text{F1 Score} = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}}$$

Precisão = 0,60

Recall = 0,75

$$\text{F1 Score} = 2 \times \frac{0,6 \times 0,75}{0,6 + 0,75} = 2 \times \frac{0,45}{1,35} = 0,6667 \text{ (ou } 66,67\%)$$

O F1 Score indica que levando em consideração tanto a precisão quanto o recall o modelo tem uma taxa de 66,67%.

Resumo das Métricas

Acurácia 98,5% - Qual a proporção total de classificações corretas (fraudes e não fraudes) em relação ao total de transações analisadas?

Precisão (Precision) 60% - Dentro do universo das fraudes que o modelo classificou como fraudes, quantas realmente eram fraudes?

Recall 75% - Dentro do universo de todas as fraudes existentes, quantas o modelo conseguiu identificar corretamente como fraudes?

F1 Score 66,67% - Como está o equilíbrio entre a precisão e o recall do modelo ao identificar fraudes? O modelo é eficaz em encontrar fraudes e também em não marcar transações normais como fraudes?

Como interpretar?

Acurácia: 98,5% de acurácia pode parecer impressionante, **mas devido ao desbalanceamento da classe** (apenas 2% das transações são fraudes), essa métrica pode ser enganosa. O modelo pode estar prevendo "não fraude" com muita frequência, pois isso representa a maioria das transações.

Precisão: Com uma precisão de 60%, isso significa que, das transações que o modelo classificou como fraudes, **apenas 60% realmente eram fraudes**. Isso sugere que o modelo está gerando um número considerável de falsos positivos (transações normais que foram incorretamente classificadas como fraudes).

Recall: Um recall de 75% indica que o **modelo conseguiu identificar 75% das fraudes reais**. Isso é um desempenho razoável, pois significa que o modelo é eficaz em detectar a maioria das fraudes existentes, mas ainda deixa 25% das fraudes não detectadas (falsos negativos).

F1 Score: O F1 Score de 66,67% reflete um equilíbrio entre a precisão e o recall. Embora não seja excelente, sugere que o modelo tem um desempenho moderado, considerando a importância tanto de detectar fraudes (recall) quanto de não acusar transações normais como fraudes (precisão).