

## REGRESSÃO LOGÍSTICA - FATORES DE RISCO DO CÂNCER DE COLO DE ÚTERO

- Ao analisar o dataset coletado do Hospital universitário de Caracas, onde visa prever indicadores de câncer de colo de útero, treinamos e testamos o modelo. Ao comparar os valores reais e previstos, obtivemos uma matriz de confusão, que possibilita distinguir e classificar da seguinte forma:

157 (VN)*	4 (FP)*
4 (FN)*	7 (VP)*

\*VN = Verdadeiro Negativo, exibe os valores reais negativos que são iguais aos valores previstos negativos.

\*FN = Falso Negativo, é um erro em que consta os valores reais falsamente previstos, quando os valores reais são positivos e os valores previstos são negativos.

\*FP = Falso Positivo, é outro erro em que consta os valores reais falsamente previstos, mas, dessa vez, ocorre quando os valores reais são negativos e os valores previstos são positivos.

\*VP = Verdadeiro Positivo, mostra os valores reais positivos que são iguais aos valores previstos positivos.

- Após esse processo, métricas são necessárias para avaliar o modelo da melhor forma possível. Tais métricas têm diferentes funções e são usadas para casos específicos. No caso do dataset utilizado, as seguintes métricas foram aplicadas:

$$1. \text{ Acurácia} = \frac{VP + VN}{VN + FN + FP + VP}$$

➡ Mais útil em dados com a mesma proporção de exemplos para a classe real e a classe prevista.

➡ O resultado da acurácia para o modelo em questão gerou o valor: 0.9534883720930233.

$$2. \text{ Precisão} = \frac{VP}{VP + FP}$$

➡ Complementa a acurácia diminuindo ao máximo o número de falsos positivos.

➡ O resultado da precisão para o modelo em questão gerou o valor: 0.6363636363636364.

3. **Recall =** 
$$\frac{VP}{VP + FN}$$

➡ Complementa a acurácia diminuindo ao máximo o número de falsos negativos.

➡ O resultado do recall para o modelo em questão gerou o valor: 0.6363636363636364.

4. **F1 - Score =** 
$$\frac{2}{\frac{1}{\text{precisão}} + \frac{1}{\text{recall}}}$$

➡ Média harmônica entre precisão e recall, é útil quando há um dataset com classes desproporcionais.

➡ O resultado do f1 - score para o modelo em questão gerou o valor: 0.6363636363636364.