

# Modelos de Machine Learning

## Classificação

De forma geral, um modelo preditivo de classificação tem como princípio decidir caso acontecerá ou não tal ação, normalmente envolve problemas de solução binária ou para identificação de imagem.

## Matriz de confusão

Considerado um dos pilares mais importantes do modelo de classificação, a matriz de confusão é responsável por "separar" as respostas apresentadas por tal modelo.

Ela é formada por uma coluna respectiva ao real valor esperado e uma linha respectiva ao valor previsto pelo modelo, fazendo a comparação entre ambos é possível retornar os valores corretos e errados previstos pelo modelo.

		Predito	
		0	1
Verdadeiro / Real	0	Verdadeiro Negativo	Falso Positivo
	1	Falso Negativo	Verdadeiro Positivo

Ao possuir a matriz de confusão completa, é possível a extração de alguns dados importantes para a análise do modelo, a fórmula abaixo representa a acurácia do modelo, sendo possível ao considerar todos os dados presentes.

$$\frac{Acertos}{Total} = \frac{VN + VP}{VN + FN + VP + FP}$$

Podemos encontrar também a precisão do modelo treinado e a sua sensibilidade, considerando o acerto ao comparar com as colunas e linhas, respectivamente.

$$\frac{VP}{VP + FP} \quad \frac{VP}{VP + FN}$$

Como estamos considerando um problema de classificação, obtemos cálculos parecidos com o anterior para considerar outros pontos, mas utilizando do valor

verdadeiro negativo. Para este tipo de problema temos o cálculo da especificidade, obtido pela fórmula abaixo.

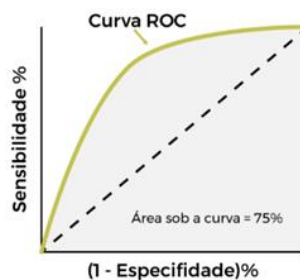
$$\frac{VN}{VN + FP}$$

## Curva ROC

A curva ROC é uma ferramenta de grande importância na avaliação da performance dos modelos, ela demonstra a relação entre a sensibilidade e a especificidade.

O gráfico referente a ele é gerado plotando a sensibilidade no eixo Y e a especificidade no eixo X. O seu intuito para apresentar um gráfico perfeito é aproximar o máximo possível essa curva ao canto superior esquerdo, se afastando de ser uma reta completa.

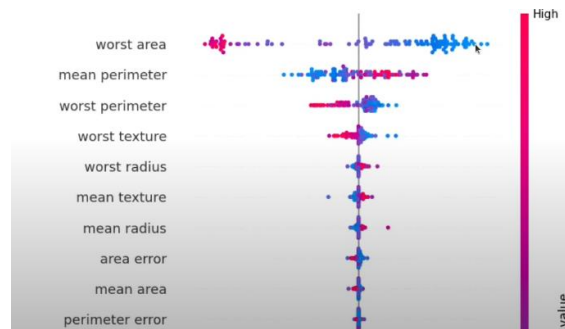
Esse modelo também possui uma utilidade para comparar diversos modelos e a eficácia presente em cada um, ajudando a escolher o mais eficaz para o problema em questão.



## SHAP

A biblioteca SHAP possui diversos mecanismos para facilitar a compreensão do que exatamente o modelo está fazendo, durante o segundo vídeo foi possível compreender melhor sobre quais parâmetros uma rede de classificação utiliza para chegar na resposta final.

Esse mecanismo pode ser aplicado para cada valor individual ou até mesmo para um gráfico de forma a mostrar todos os parâmetros.

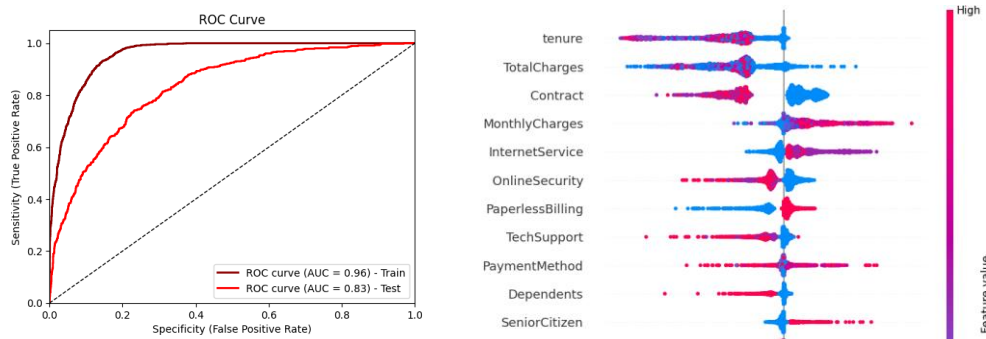


## Atividade criada

A atividade feita apresenta uma criação de um modelo utilizando a biblioteca RandomForest para decidir caso o cliente continuará ou não a utilizar do produto.

Utilizando um conjunto de dados [Telco Customer Churn](#), foi possível treinar com aproximadamente 7 mil dados, tornando um modelo simples, porém com uma boa eficiência. A precisão foi de aproximadamente 79%, mas com possíveis pontos a serem aprimorados.

O código aplica conceito de ambos os vídeos ao gerar e mostrar sua curva ROC e a apresentação gráfica utilizando o SHAP para apresentar quais os parâmetros mais importantes.



## Conclusão

Este card abordou os principais conceitos sobre modelos de classificação, destacando a importância da matriz de confusão e da curva ROC na avaliação de performance. Além disso, foi discutido o uso da biblioteca SHAP para facilitar a interpretação dos modelos de machine learning. Essas ferramentas fornecem uma base sólida para analisar e melhorar modelos preditivos, garantindo maior compreensão e eficácia nas previsões.