

Big Data Analytics com R e Microsoft Azure Machine Learning



Data Science Academy



Classificação

- Aprendizagem Supervisionada
- Classe de modelos para categorizar valores (utiliza variáveis categóricas)
- Métodos Two-class e Multi-class
- Erros são medidos pelas taxas de classificações incorretas
- Alguns erros podem ser mais críticos que outros e trade-off terão que ser feitos



Data Science Academy



Performance dos Modelos de Classificação

Confusion Matrix

Atuais	Previstos	
	Sim	Não
Sim	True Positive (TP)	False Negative (FN)
Não	False Positive (FP)	True Negative (TN)



Data Science Academy



Performance dos Modelos de Classificação

Medidas de Performance

Medida de Performance	Definição
Accuracy	Total de resultados corretos / total de casos analisados
Recall	Total de resultados positivos / Total de resultados corretos
Precision	Proporção de "true" / Total de resultados corretos
F-Score	$F = 2 * TP / (2 * TP + FP + FN)$ - Balanceamento entre Precision e Recall
AUC	AUC = Area Under the Curve. Plot de TP no eixo y e FP no eixo x



Data Science Academy

Performance dos Modelos de Classificação

Medidas de Performance

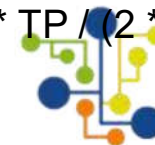
	Previstos	
Atuais	Sim	Não
Sim	True Positive (TP)	False Negative (FN)
Não	False Positive (FP)	True Negative (TN)

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

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

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{F-Score} = 2 * TP / (2 * TP + FP + FN)$$



Data Science Academy