# Ciência de Dados (Big Data Processing and Analytics)

Big Data Analytics – Mineração e Análise de Dados

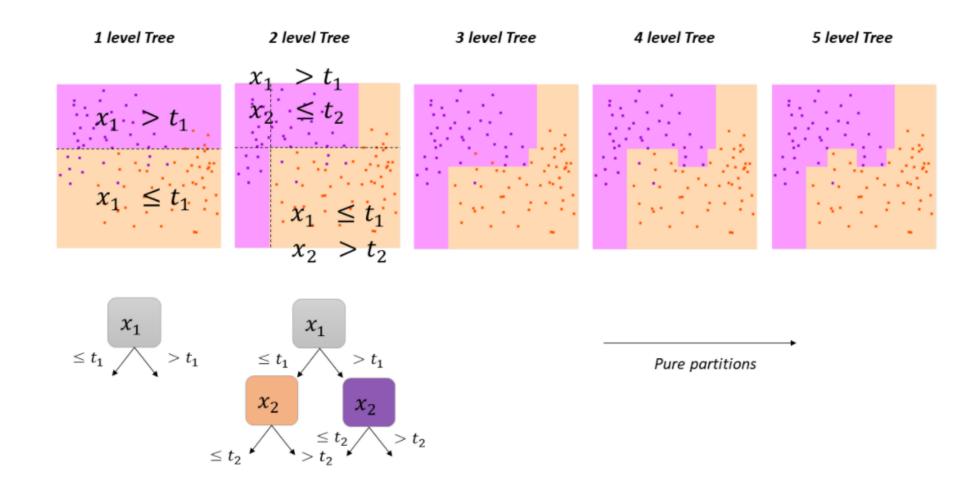




### TRILHA 5 Árvores de Decisão, Seleção de Atributos e outros Classificadores

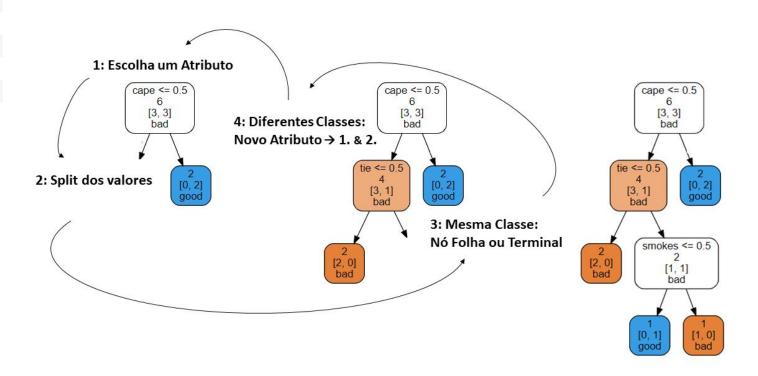
Parte A

### Árvores de Decisão: Um Modelo Partitivo



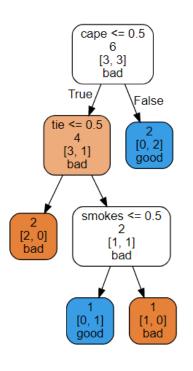
### Árvores de Decisão: Hunt

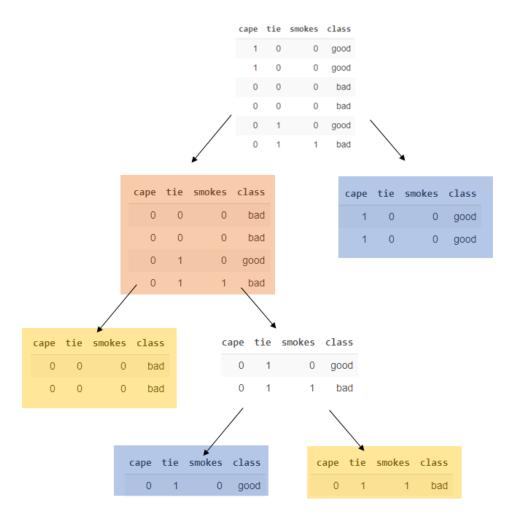
	Name	sex	smokes	tie	mask	cape	ears	class
0	Batman	male	no	no	yes	yes	yes	good
1	Robin	male	no	no	yes	yes	yes	good
2	Catwoman	female	no	no	yes	no	yes	bad
3	Joker	male	no	no	no	no	no	bad
4	Alfred	male	no	yes	no	no	no	good
5	Penguin	male	yes	yes	no	no	no	bad



### Árvores de Decisão: Hunt

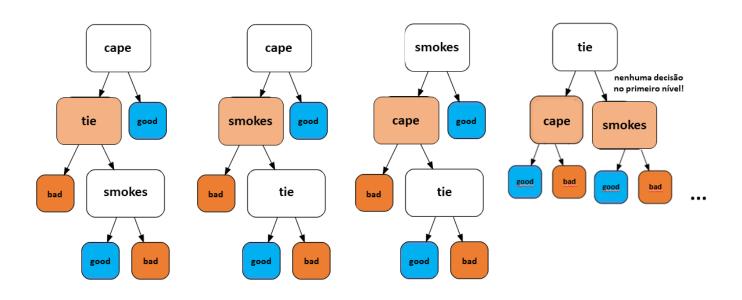
	Name	sex	smokes	tie	mask	cape	ears	class
0	Batman	male	no	no	yes	yes	yes	good
1	Robin	male	no	no	yes	yes	yes	good
2	Catwoman	female	no	no	yes	no	yes	bad
3	Joker	male	no	no	no	no	no	bad
4	Alfred	male	no	yes	no	no	no	good
5	Penguin	male	yes	yes	no	no	no	bad



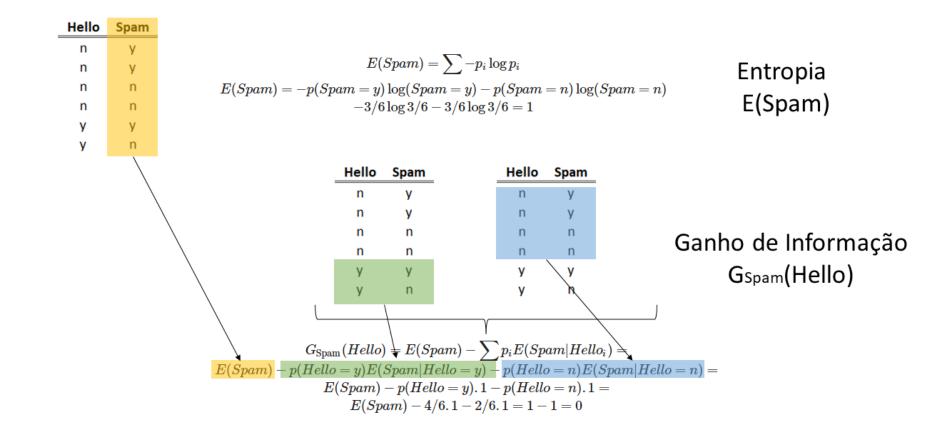


## Árvores de Decisão: Qual a melhor Árvore?

	Name	sex	smokes	tie	mask	cape	ears	class
0	Batman	male	no	no	yes	yes	yes	good
1	Robin	male	no	no	yes	yes	yes	good
2	Catwoman	female	no	no	yes	no	yes	bad
3	Joker	male	no	no	no	no	no	bad
4	Alfred	male	no	yes	no	no	no	good
5	Penguin	male	yes	yes	no	no	no	bad



# Árvores de Decisão: Entropia e Ganho de Informação

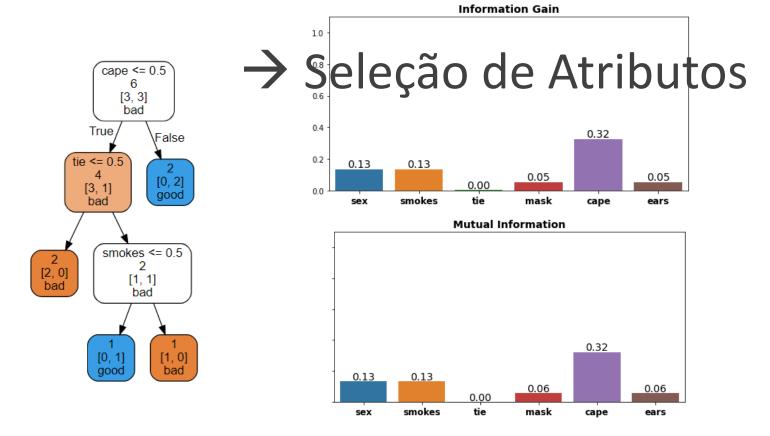


### TRILHA 5 Árvores de Decisão, Seleção de Atributos e outros Classificadores

Parte B

### Ganho de Informação = Informação Mútua

	Name	sex	smokes	tie	mask	cape	ears	class
0	Batman	male	no	no	yes	yes	yes	good
1	Robin	male	no	no	yes	yes	yes	good
2	Catwoman	female	no	no	yes	no	yes	bad
3	Joker	male	no	no	no	no	no	bad
4	Alfred	male	no	yes	no	no	no	good
5	Penguin	male	yes	yes	no	no	no	bad

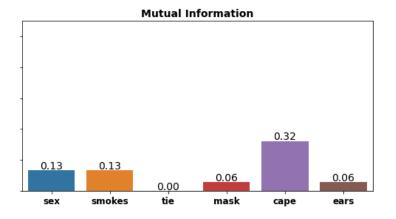


### Seleção de Atributos SelectKBest

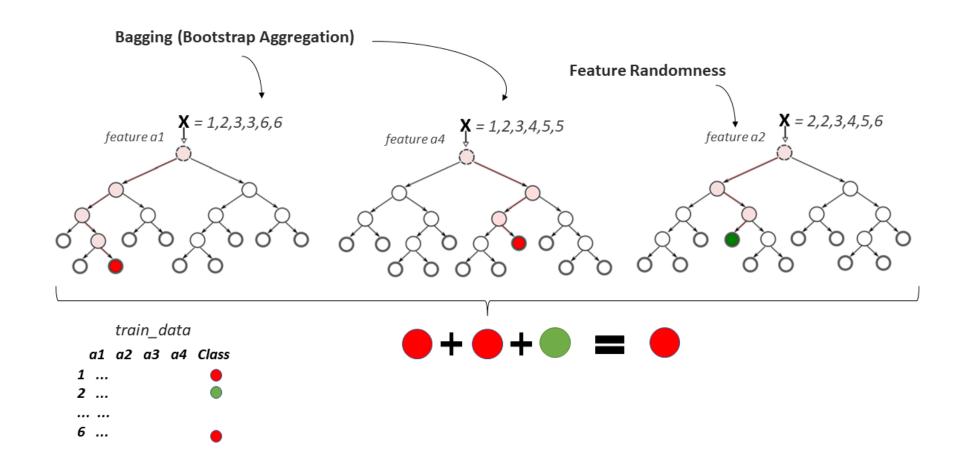
	Name	sex	smokes	tie	mask	cape	ears	class
0	Batman	male	no	no	yes	yes	yes	good
1	Robin	male	no	no	yes	yes	yes	good
2	Catwoman	female	no	no	yes	no	yes	bad
3	Joker	male	no	no	no	no	no	bad
4	Alfred	male	no	yes	no	no	no	good
5	Penguin	male	yes	yes	no	no	no	bad







#### **Ensemble Models: Random Forest**



#### **Selecionando Modelos**

