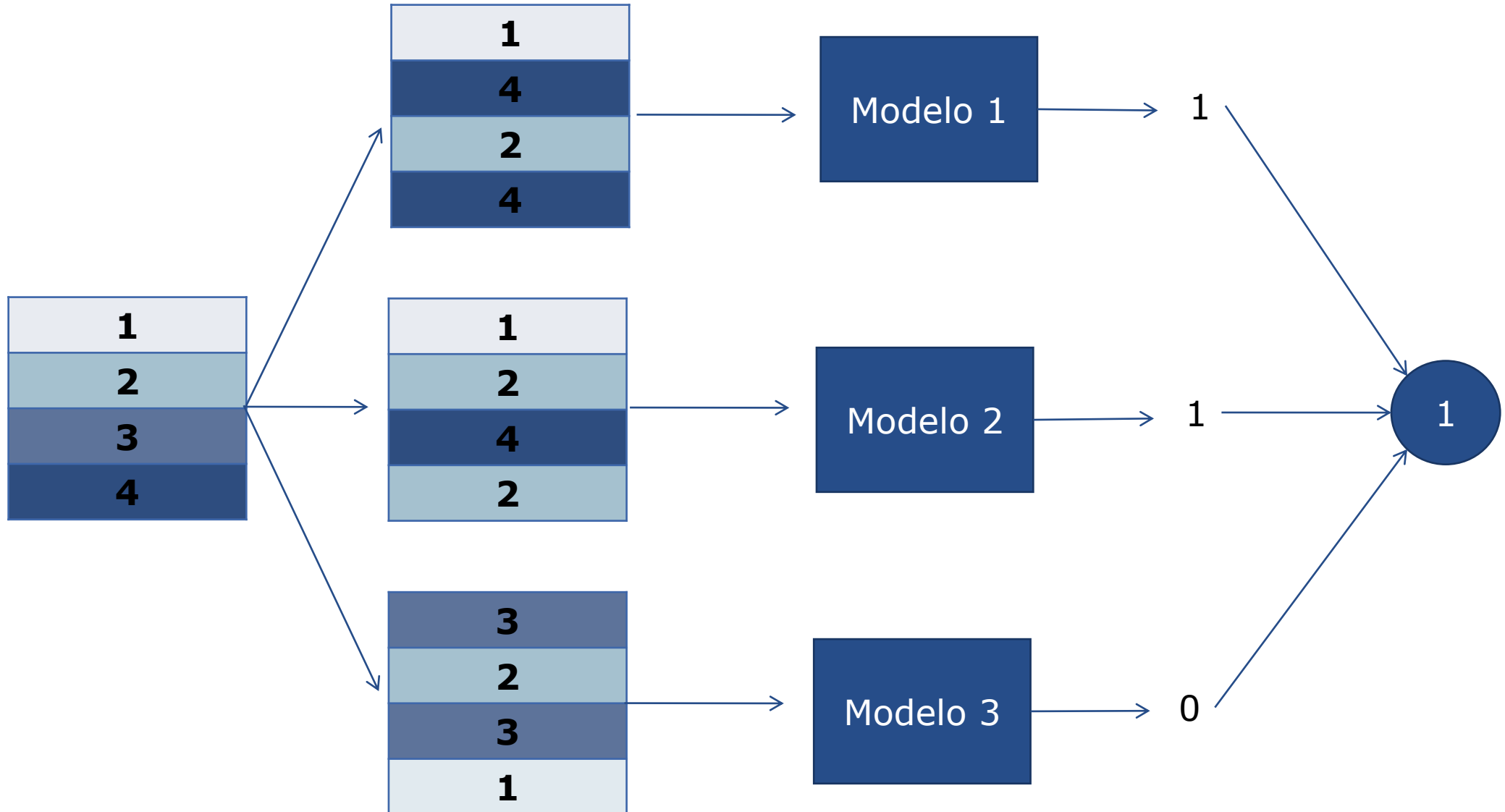




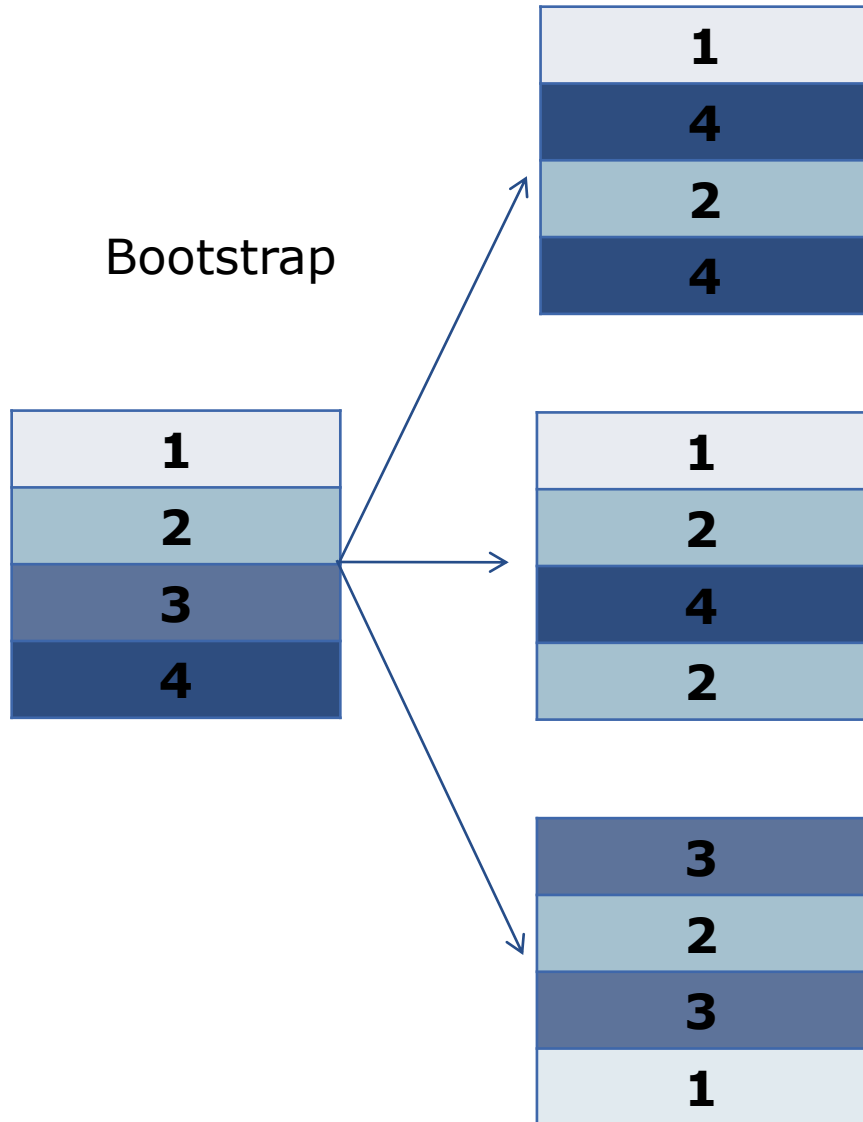
escola  
britânica de  
artes criativas  
& tecnologia

# **Profissão Cientista de Dados** **Random Forest**

# Bagging - Bootstrap Aggregating



# Random Forest



# Random Forest

Bootstrap + Feature selection

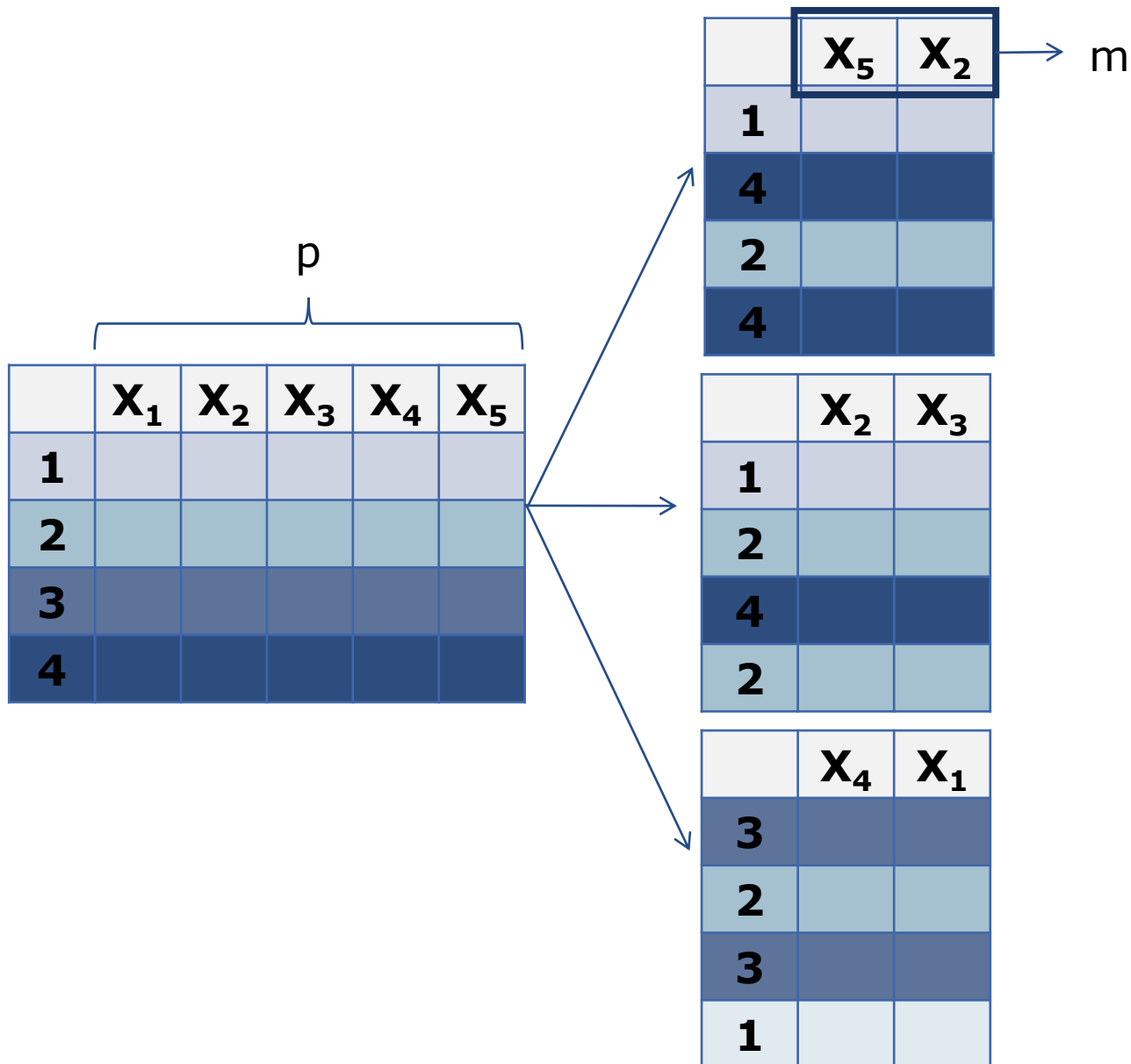
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
1					
2					
3					
4					

	$X_5$	$X_2$
1		
4		
2		
4		

	$X_2$	$X_3$
1		
2		
4		
2		

	$X_4$	$X_1$
3		
2		
3		
1		

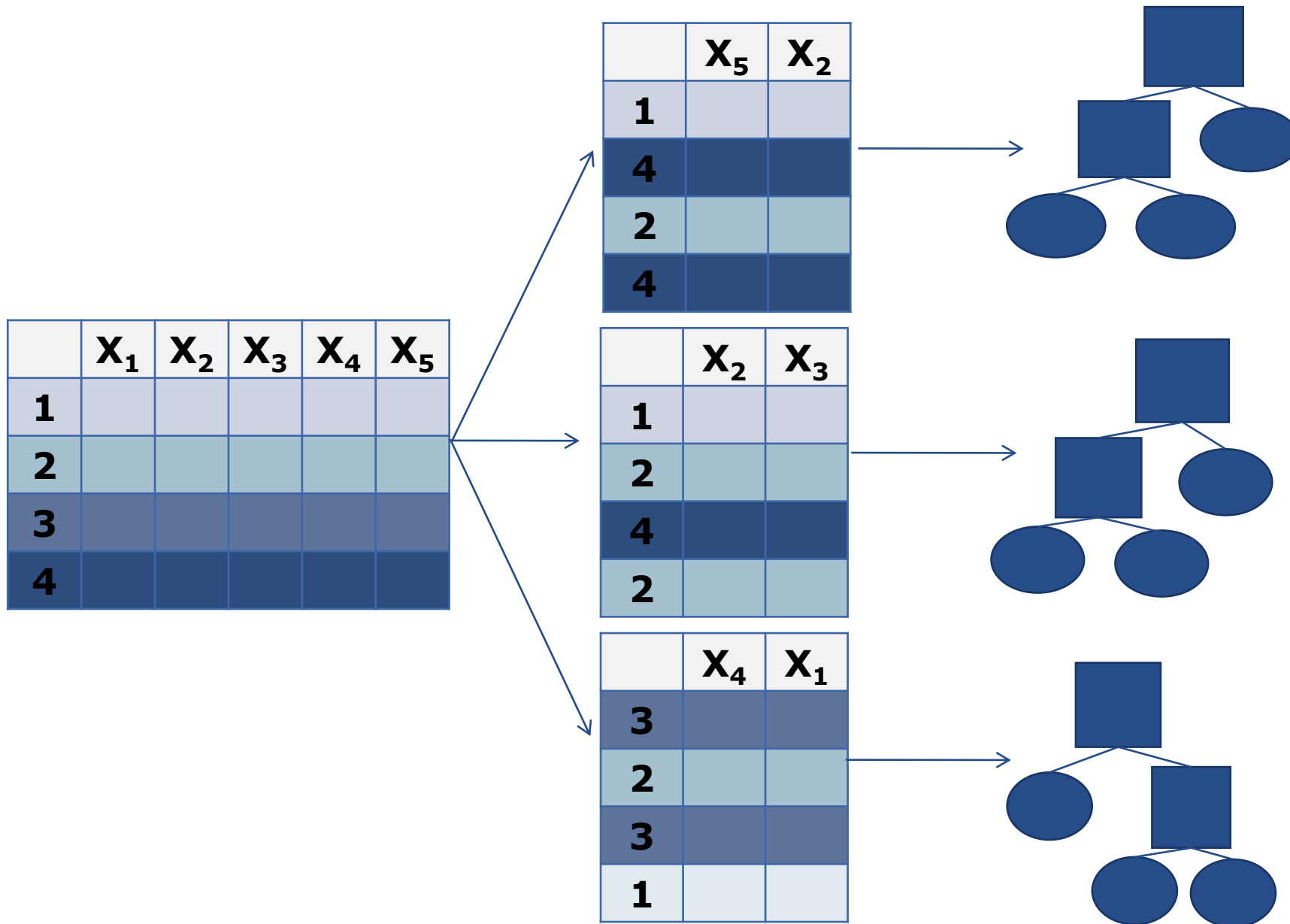
# Random Forest



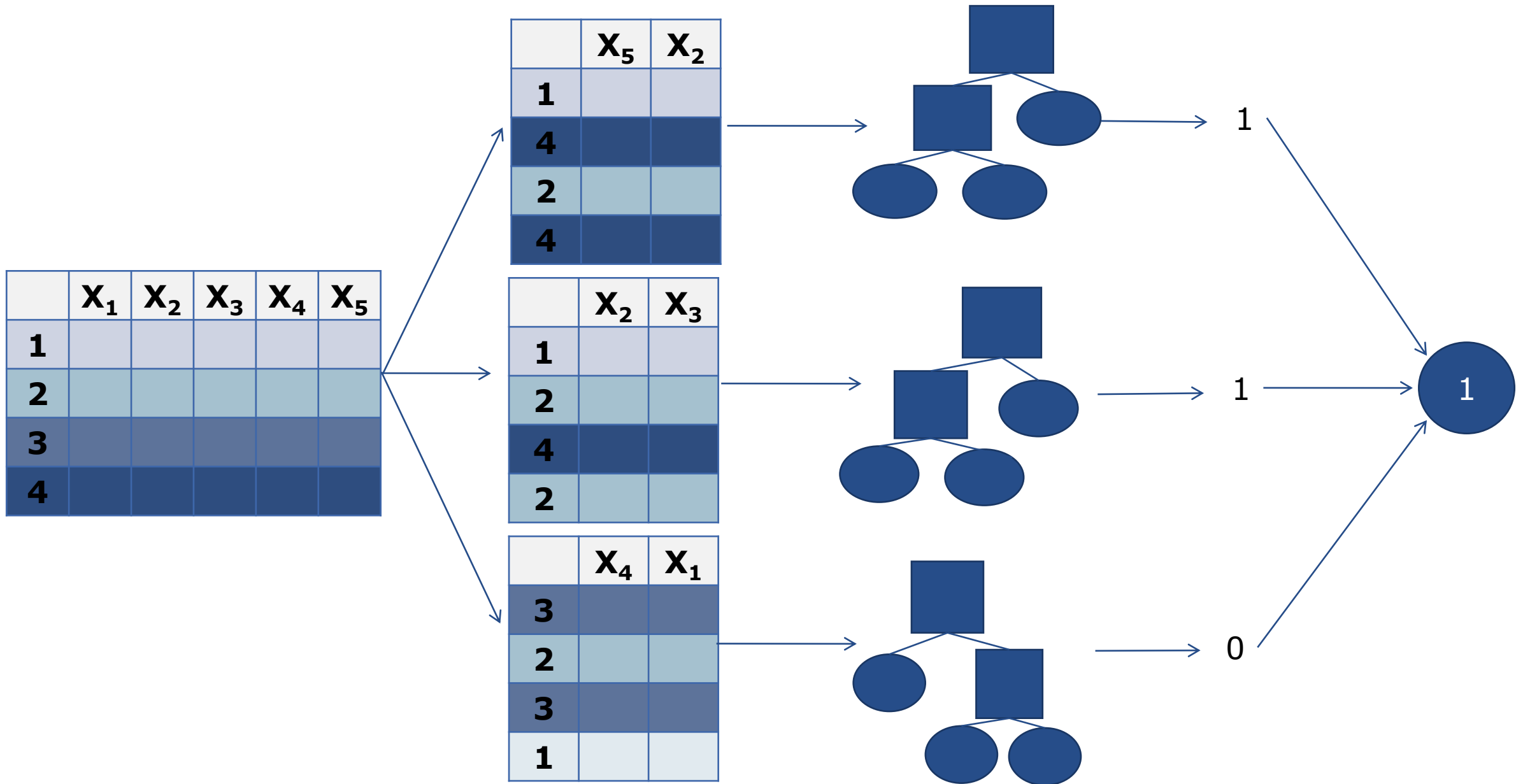
Segundo Leo Breiman (2001), devemos usar as seguintes condições como padrão:

- Para classificação  $\rightarrow m = \sqrt{p}$
- Para regressão  $\rightarrow m = p/3$

# Random Forest



# Random Forest



# Random Forest



Sabedoria das multidões

O conhecimento coletivo de uma multidão diversificada e independente de pessoas normalmente excede o conhecimento de qualquer indivíduo único e pode ser usado através do voto para se chegar a uma decisão. (Surowiecki, 2004).

Random forest funciona melhor que o Bagging, pois as amostradas são mais independentes (menor correlação).



Robusto a overfitting

É robusto a overfitt, mas há casos em que podem ocasionar o overfitt.