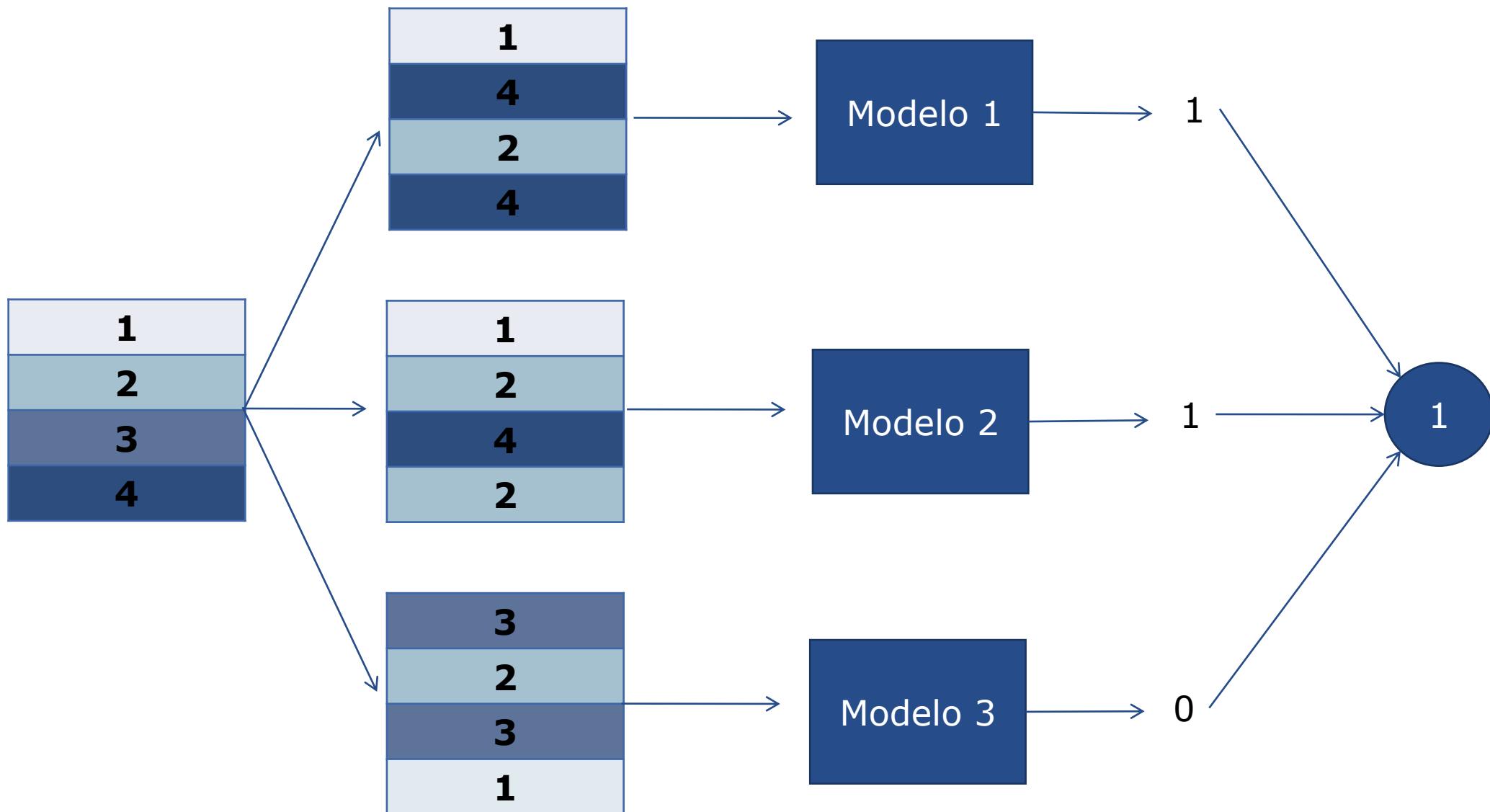




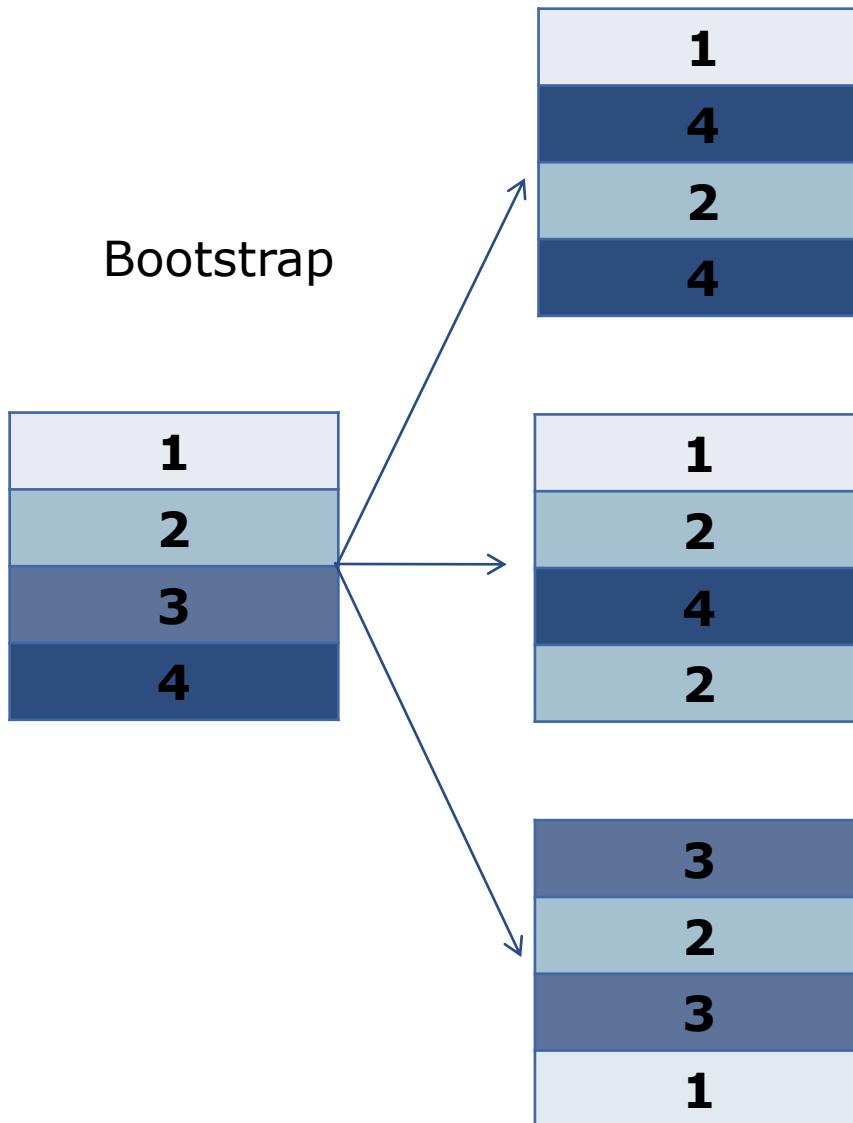
escola
britânica de
artes criativas
& tecnologia

**Profissão Cientista de Dados
Random Forest**

Bagging - Bootstrap Aggregating



Random Forest



Random Forest

Bootstrap + Feature selection

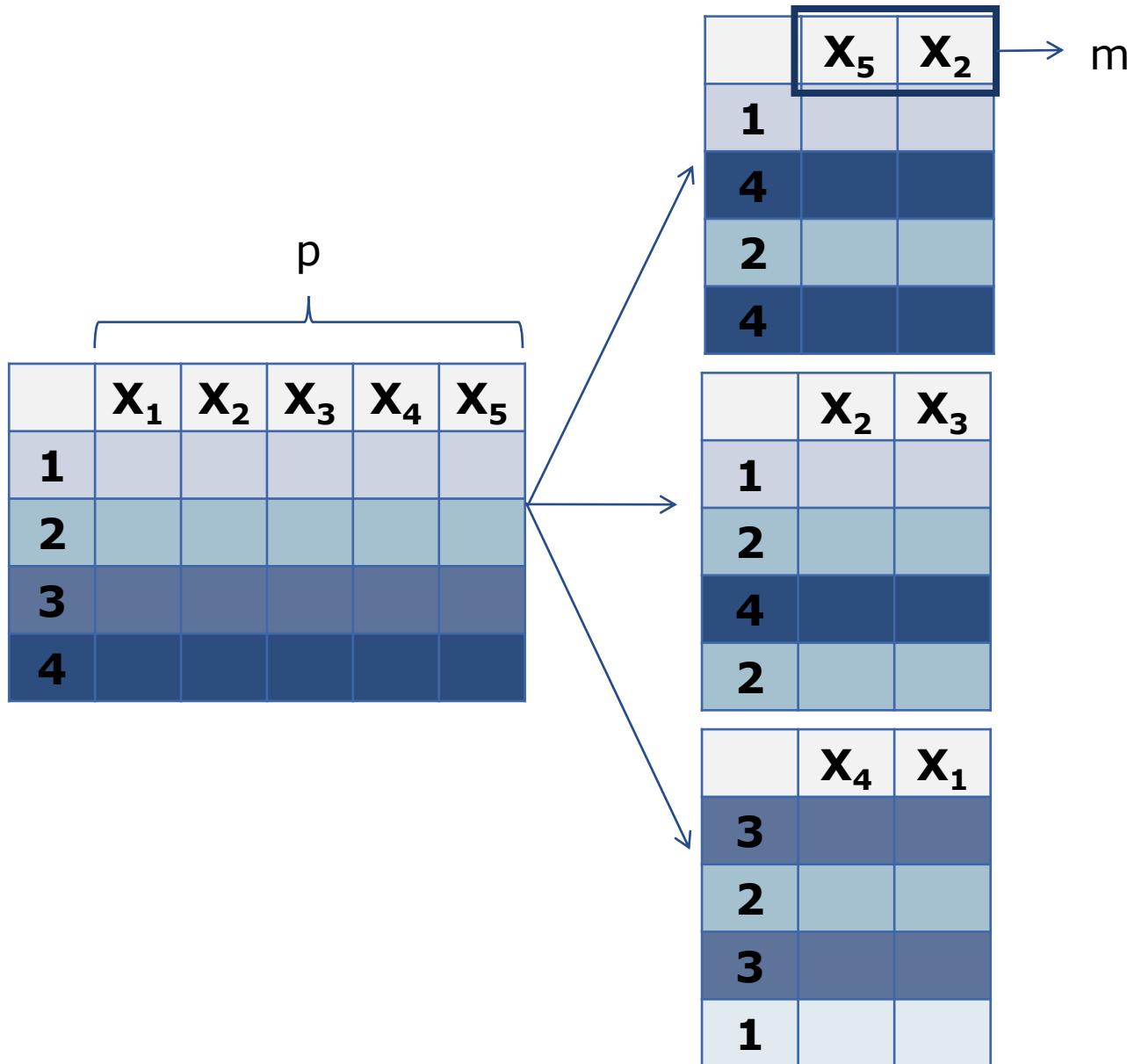
| | \mathbf{X}_1 | \mathbf{X}_2 | \mathbf{X}_3 | \mathbf{X}_4 | \mathbf{X}_5 |
|---|----------------|----------------|----------------|----------------|----------------|
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |

| | \mathbf{X}_5 | \mathbf{X}_2 |
|---|----------------|----------------|
| 1 | | |
| 4 | | |
| 2 | | |
| 4 | | |

| | \mathbf{X}_2 | \mathbf{X}_3 |
|---|----------------|----------------|
| 1 | | |
| 2 | | |
| 4 | | |
| 2 | | |

| | \mathbf{X}_4 | \mathbf{X}_1 |
|---|----------------|----------------|
| 3 | | |
| 2 | | |
| 3 | | |
| 1 | | |

Random Forest



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

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

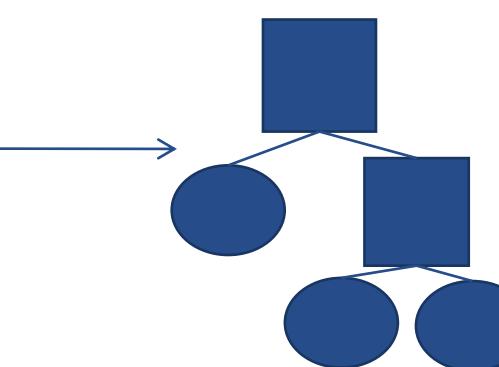
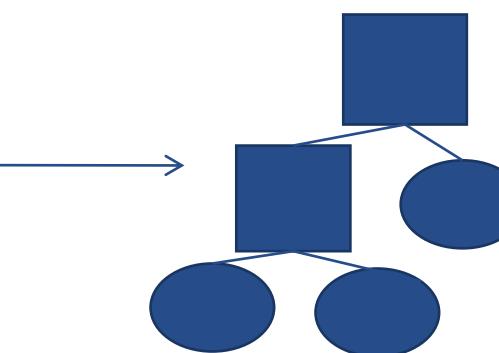
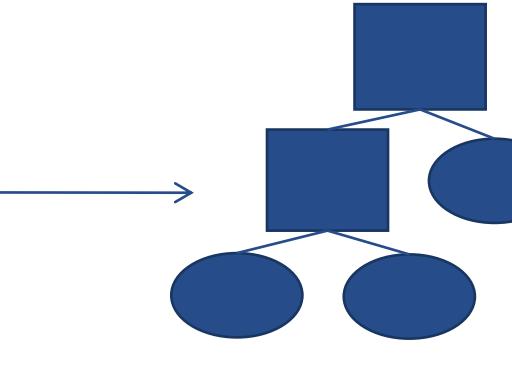
Random Forest

| | X_1 | X_2 | X_3 | X_4 | X_5 |
|---|-------|-------|-------|-------|-------|
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |

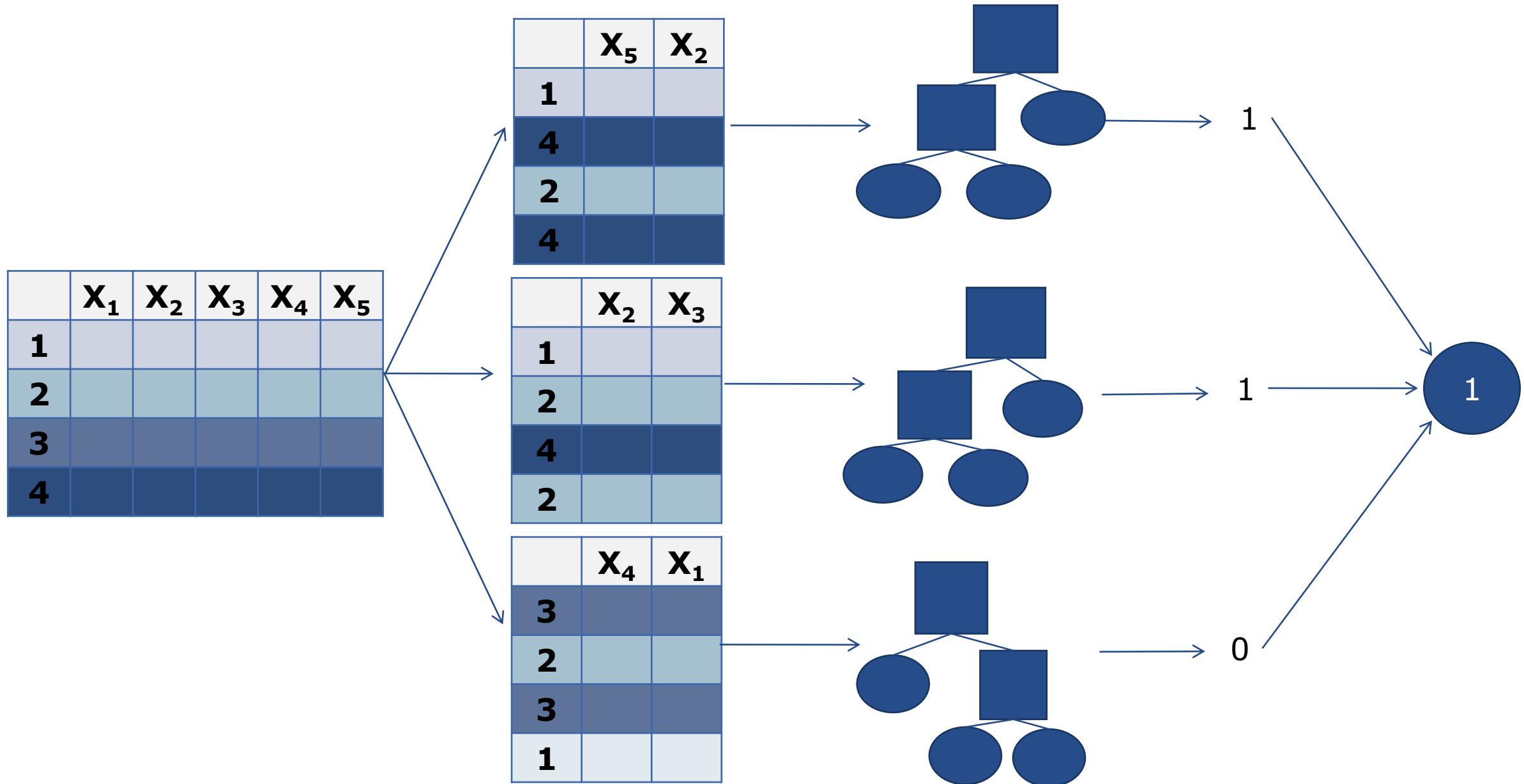
| | X_5 | X_2 |
|---|-------|-------|
| 1 | | |
| 4 | | |
| 2 | | |
| 4 | | |

| | X_2 | X_3 |
|---|-------|-------|
| 1 | | |
| 2 | | |
| 4 | | |
| 2 | | |

| | X_4 | X_1 |
|---|-------|-------|
| 3 | | |
| 2 | | |
| 3 | | |
| 1 | | |



Random Forest



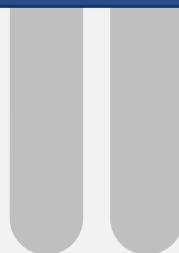
Random Forest



Sabedoria das multidões

O conhecimento coletivo de uma multidão diversificado 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. (Groseclose & Garrowiecki, 2004).

Random forest funciona melhor que o Bagging, já que os árvores amostradas são mais independentes (menor correlação entre os resultados).



Robusto a overfitting

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