

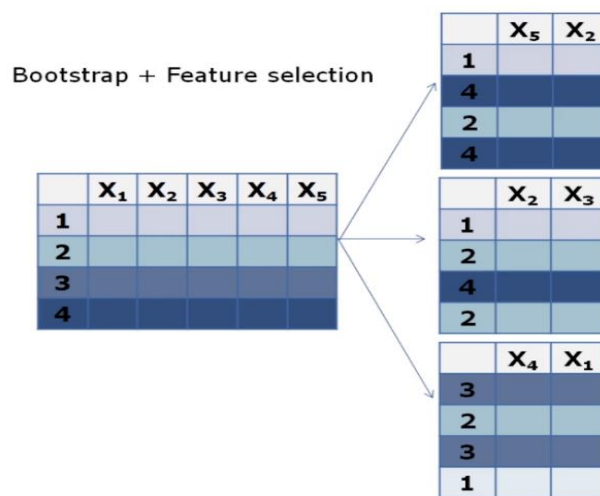
Random Forest

O Random Forest, que em português significa **Floresta Aleatória**, é um algoritmo que irá criar muitas árvores de decisão, de maneira aleatória, formando uma “floresta”, onde cada árvore será utilizada na escolha do resultado final, em uma espécie de votação.

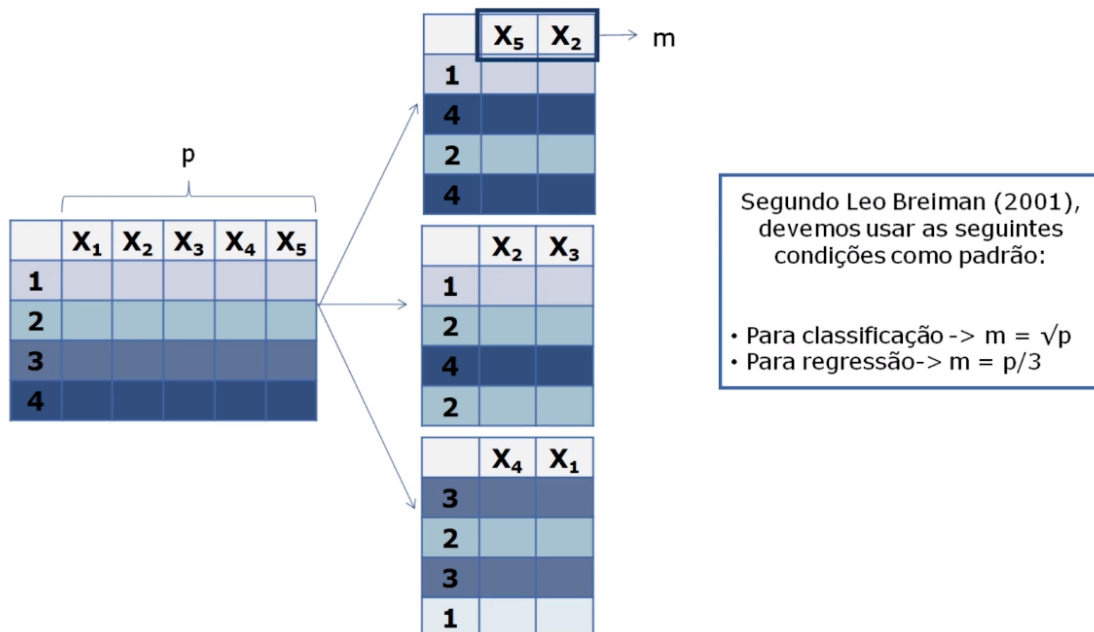
Para aprofundar melhor, vamos observar suas etapas:

- Seleção de amostras:

Agora iremos trabalhar com a seleção de linhas e colunas, sendo a primeira utilizando amostragem com repetição (Bootstrap) e a segunda determinando um valor para a quantidade com seleção aleatória (Feature Selection). Como resultado teremos os subconjuntos de dados, como ilustrado abaixo.

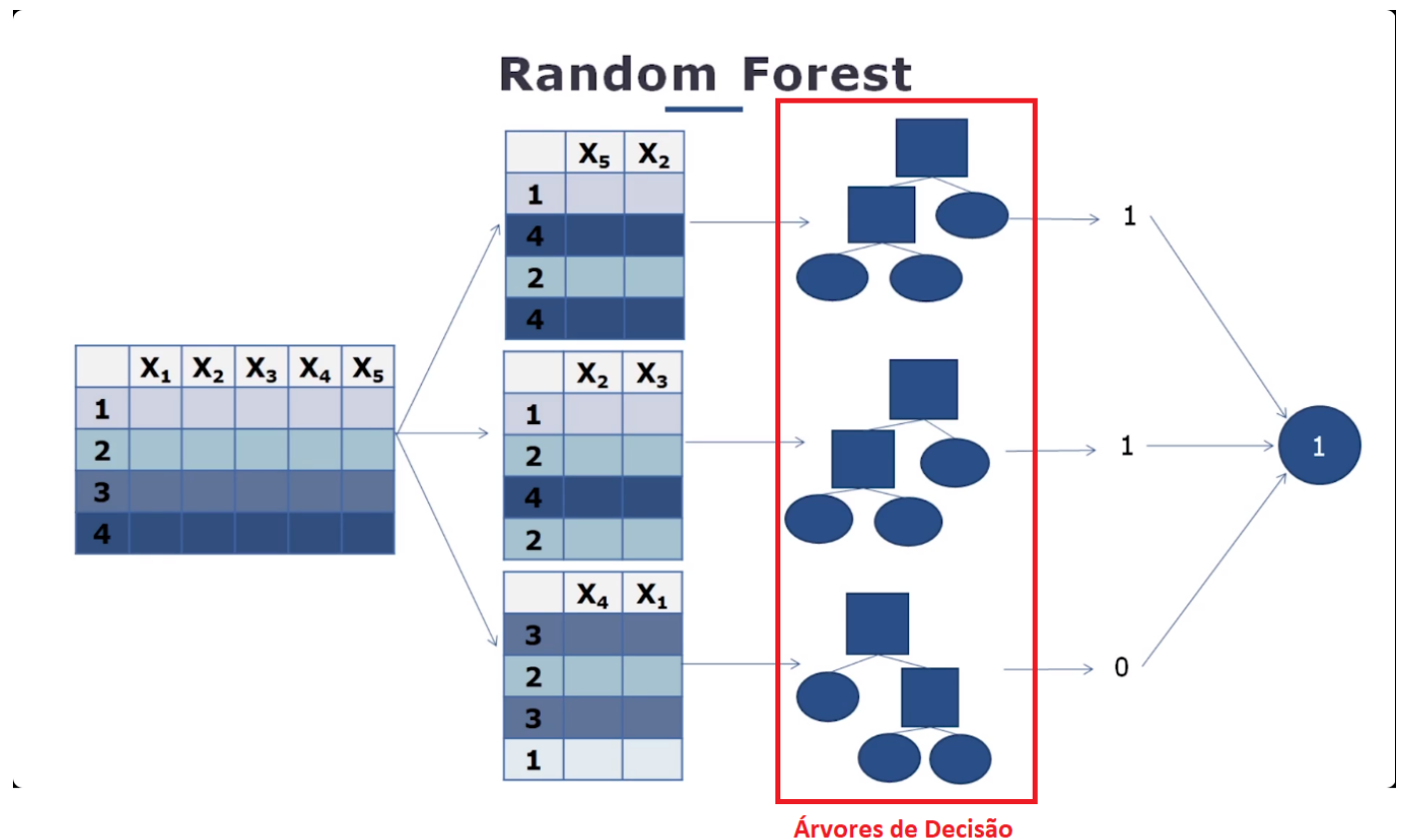


Obs.: A determinação da quantidade de colunas selecionadas obedece às condições determinadas por Leo Breiman.



- Treinamento das Árvores:

Após a seleção das amostras, será feito o treinamento de uma árvore de decisão para cada subconjunto de dados e os resultados preditos serão utilizados para montar uma tabela de frequência para determinar o resultado final, com base na quantidade de votos.



Por fim, o Random Forest faz uso da “sabedoria das multidões” pois, segundo Surowiecki, “O conhecimento coletivo de um grupo diversificado e independente de pessoas normalmente excede o conhecimento de qualquer indivíduo único e pode ser usado através do voto para se tomar a decisão”. Isso torna o algoritmo robusto ao overfitting, mas não imune.