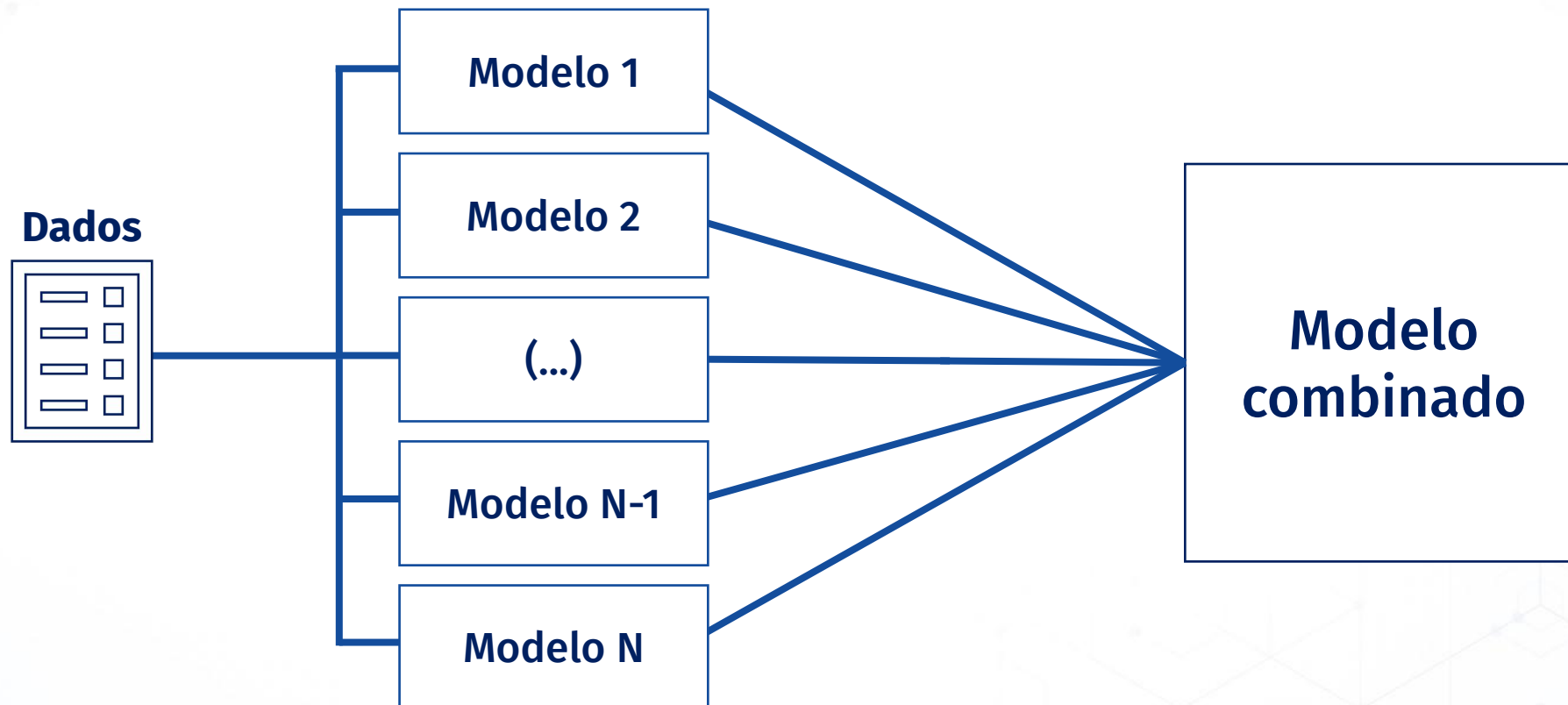
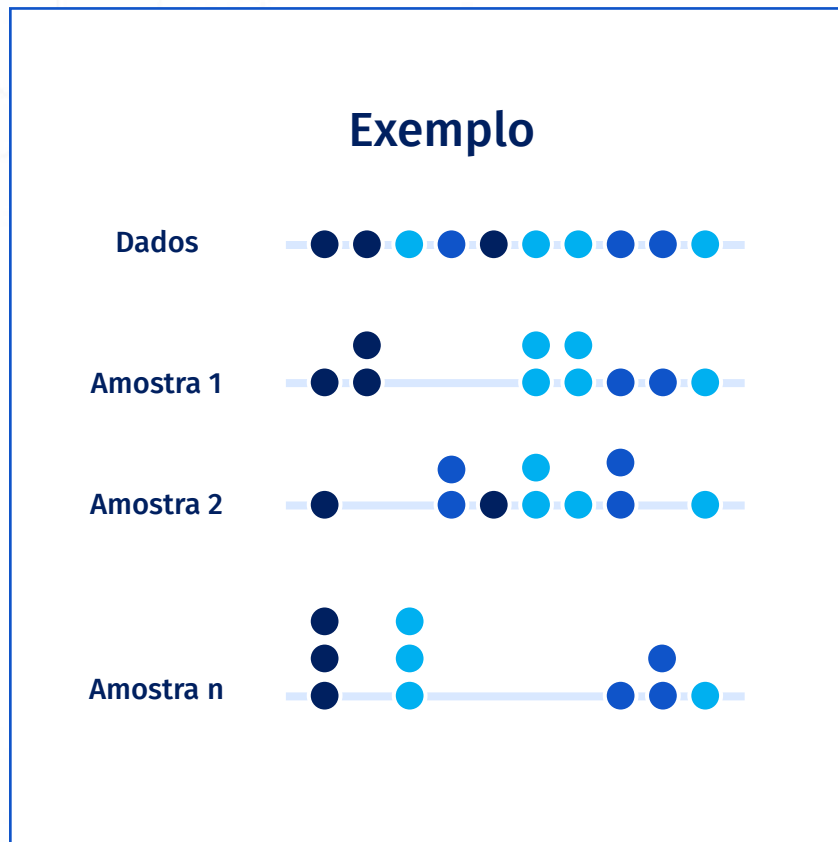


# COMBINAÇÃO DE MODELOS

Precisamos construir **modelos simples**, para então **combiná-los** e alcançar um **resultado final superior** através da combinação desses modelos



# RÉPLICAS BOOTSTRAP



São amostragens de um conjunto de dados, com reposição que em geral, são do mesmo tamanho do conjunto de dados.

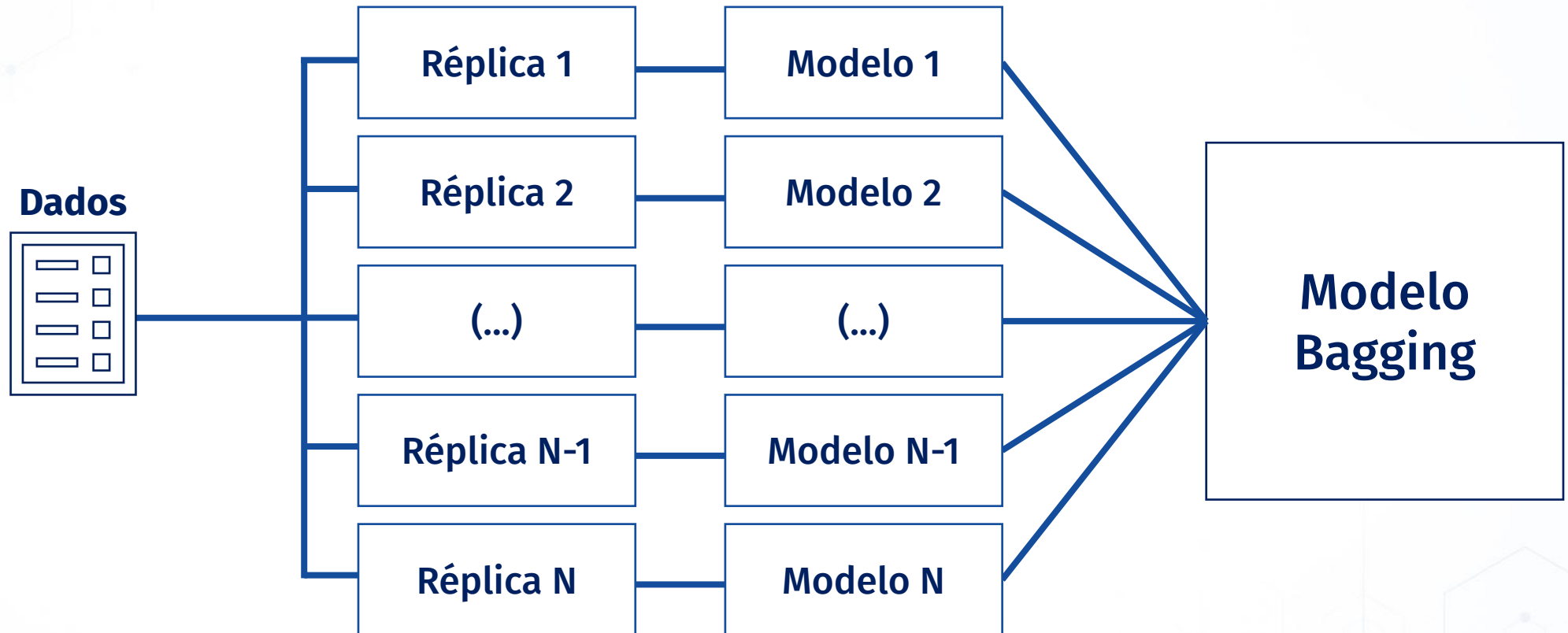
Esse procedimento induz uma média de aproximadamente 63% de observações distintas do conjunto de dados original.

É como se fossem pesos associados aos elementos de um conjunto de dados.

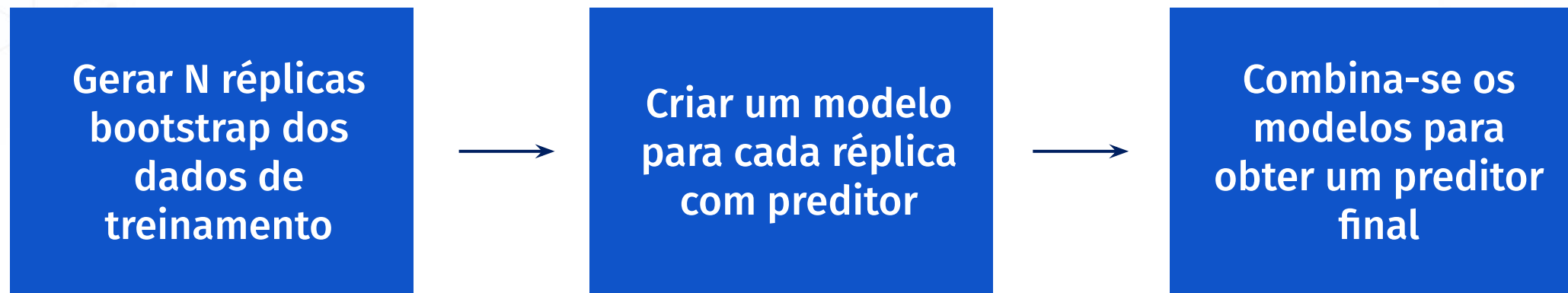
No exemplo ao lado, cada coluna formada por dois ou mais dados significa que a amostra sorteou o mesmo dado duas vezes.

# **MÉTODO BAGGING**

# MÉTODO BAGGING



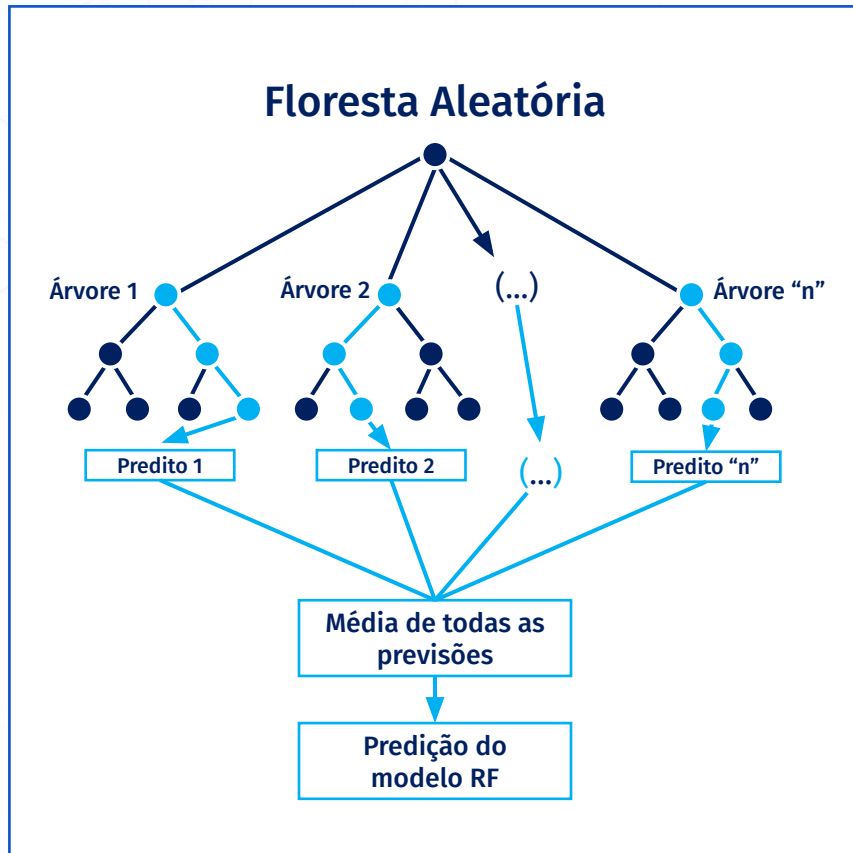
## O algoritmo bagging é



A eficiência do método está atrelada com a **diversidade nos dados**.  
Nesse algoritmo, troca-se **interpretação por acurácia**.

# **RANDOM FORESTS**





**Método:** São criadas árvores em que as variáveis utilizadas na classificação são um subconjunto de todas as variáveis disponíveis.

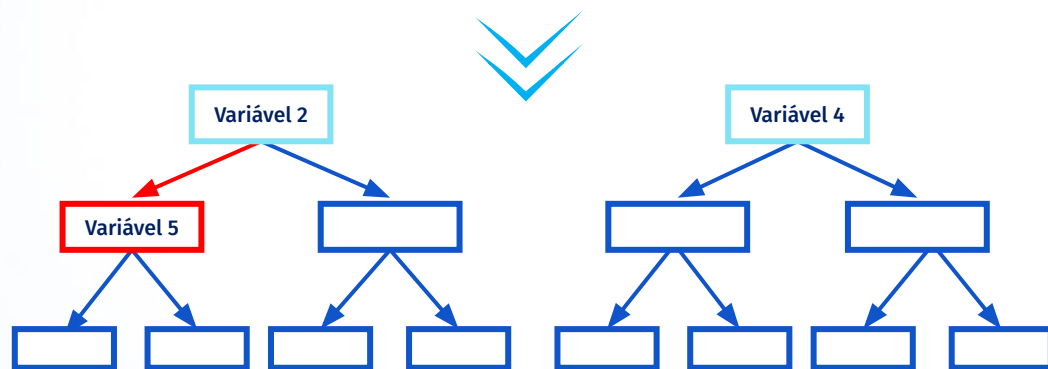
Em cada divisão de nó nas árvores geradas, um subconjunto diferente de variáveis é utilizado.

O resultado final é a **combinação da classificação apresentada por cada árvore da floresta**

É um método bastante utilizado na prática, pois tem desempenho bom na maioria dos casos

## Exemplo

	Variável 1	Variável 2	Variável 3	Variável 4	Variável 5
Amostra 1					
Amostra 3					
Amostra 4					
Amostra 6					
Amostra 6					



No Random Forest, o **algoritmo** irá **escolher de maneira aleatória** duas ou mais **variáveis** e então definirá qual será o primeiro nó.

Para escolha da variável do próximo nó, serão escolhidas pelo menos duas variáveis, **excluindo as já selecionadas anteriormente**. Desta forma a árvore será construída até o último nó.

A quantidade de **variáveis** a serem escolhidas **pode ser definida na criação do modelo**.



