

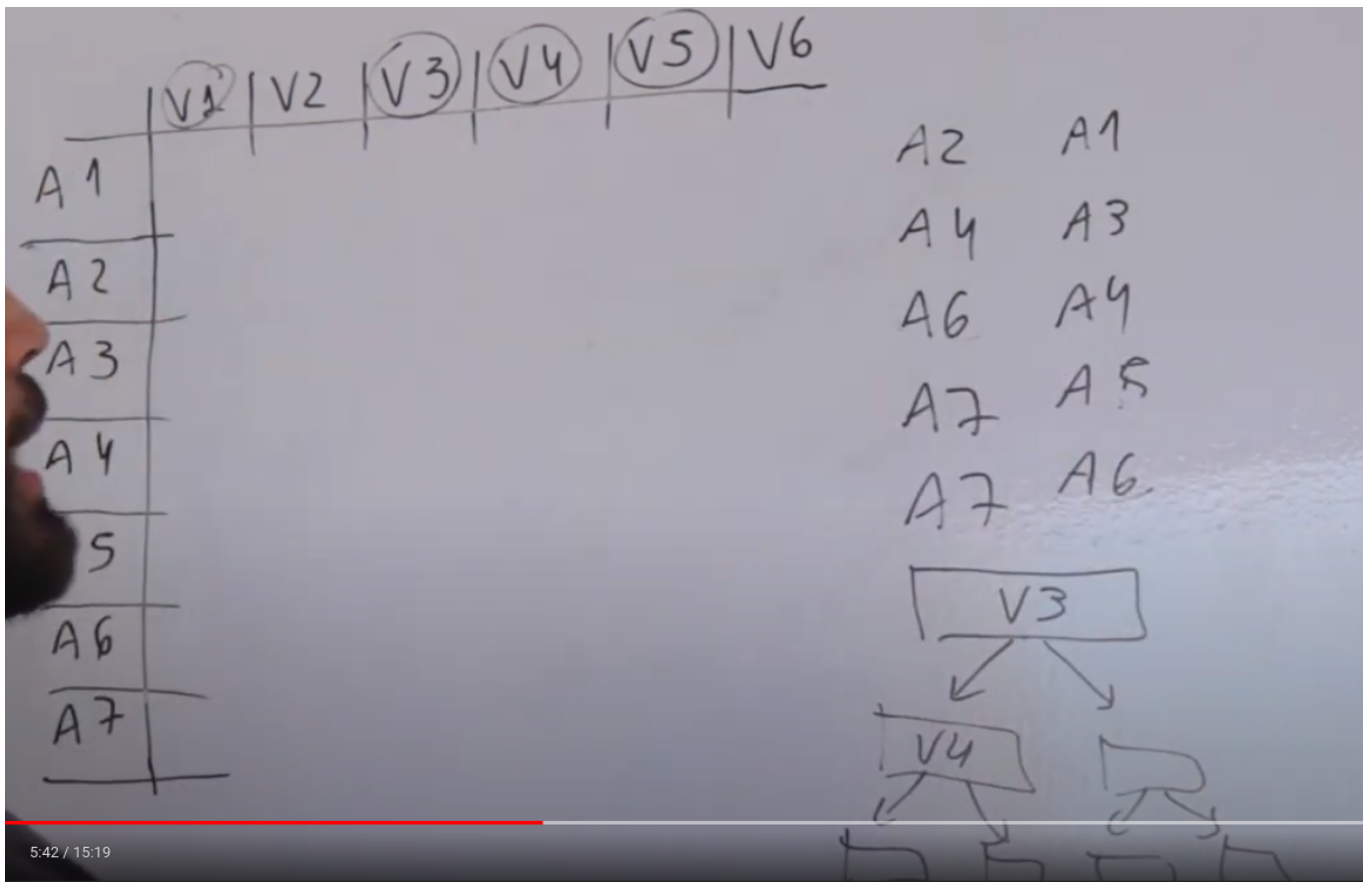
## RandomForest (Conceito)

O método RandomForest utiliza o conceito de árvores de decisão, entretanto é totalmente aleatório. Basicamente temos um número específico de árvores que o algoritmo irá construir.

Para iniciar a construção o modelo escolhe aleatoriamente um número de amostras e um número máximo de variáveis (caso não informado irá usar raiz de  $N$ , onde  $N$  é o número de variáveis), assim ele utiliza o índice Gini ou entropia para definir quais das variáveis selecionadas é a melhor para o início da árvore. Após isso ele constrói os ramos de cada árvore realizando o mesmo processo, seleciona o número de variáveis aleatória, calcula qual o melhor índice Gini ou entropia, e vai separando apenas os dados que foram aleatoriamente selecionados no início da construção da árvore.

Por fim cada árvore será submetida a uma amostra e individualmente terão o resultado, por exemplo, em um problema binário de classificação, entre 10 árvores, 7 delas disseram que é 0 e somente 3 disseram 1, portanto a classe selecionada será 0. Dessa forma temos 10 árvores distintas que agregaram para o resultado final.

Em problemas de regressão, basta extrair a média do resultado das árvores finais.



Amostras podem ser selecionadas mais de uma vez, ou seja, será feito uma re-amostragem ou bootstrap dos dados.