Extra Trees (Conceito)

Como também acontece no algoritmo RandomForest, as Extra Trees são muito mais aleatórias. Nesse caso também vamos inciar selecionando aleatoriamente um número de variáveis, porém no caso do RandomForest a escolha da melhor variável para iniciar a árvore ou o nó é decidida a partir do índice Gini ou do ganho de informação com entropia, já em Extra Trees essa escolha também vai ser randômica, escolhendo um treshold randômico, ou seja, o split aleatório. Após isso é utilizado o índice Gini ou ganho de informação para decidir os nós.

Portanto a vantagem desse algoritmo é ser menos invesado, assim como random forest, porém tornando cada árvore um pouco mais fraca, tentando aumentar a generalização do modelo para novos dados. Entretanto, como a escolha das variáveis é aleatória, podemos acabar escolhendo algumas que não são tão úteis realmente para o problema, e sim só ruído na base de dados, dado destaque grande para variáveis pouco importantes.

Obs: O algoritmo vem com bootstrap de amostras desativado, portanto ele utiliza todas as amostras para criar a árvore, diferente do RandomForest que apenas seleciona uma parte das amostras e também pode repeti-las.