Substituindo dados "Missing" com um modelo de regressão

Ao invés de utilizar a média ou mediana para substituir variáveis NA's no conjunto de dados, utilizaremos as outras variáveis para prever o valor adequado para a que não foi preenchida utilizando modelos de regressão, como se essa variável fosse a "target" e as outras "preditoras".

Nem sempre esse método é o mais eficiente, podemos ainda utilizar a média ou mediana para a substituição e ainda sim ter resultados eficientes e até melhores. E também alguns modelos são melhores do que outros em casos específicos, como por exemplo, modelos de regressão com regularização tendem a ser melhores do que árvores de decisão quando alguns pontos estão muito próximos, uma vez que a árvore não conseguiria separar adequadamente sem o devido peso às features.