



**Data Science
Academy**

www.datascienceacademy.com.br

Machine Learning

**Métodos Ensemble
Combinação de Preditores**



Métodos Ensemble permitem aumentar consideravelmente o nível de precisão nas previsões do modelo preditivo.

Quando falamos de Métodos Ensemble, estamos falando na verdade na combinação de preditores, combinação de modelos. A ideia principal dos algoritmos que utilizam método ensemble é a de utilizar a saída final de diferentes preditores para agregá-los em um único preditor. Basicamente, ensemble é uma técnica de aprendizagem supervisionada para combinar vários modelos fracos a fim de produzir um modelo forte. Um Modelo Ensemble funciona melhor, quando agrupamos modelos com baixa correlação. Um bom exemplo de como os métodos ensemble são comumente usados para resolver problemas de ciência de dados é o algoritmo Random Forest, bastante usando em seleção de variáveis.

Os dois algoritmos mais populares de combinação de preditores são Bagging e Boosting, mas existe ainda o Voting ensemble.

Usamos Bootstrap Aggregating (Bagging) para construção de múltiplos modelos (normalmente do mesmo tipo) a partir de diferentes subsets no dataset de treino.

Usamos Boosting para construção de múltiplos modelos (normalmente do mesmo tipo), onde cada modelo aprende a corrigir os erros gerados pelo modelo anterior, dentro da sequência de modelos criados.

Já o método Voting é usado na construção de múltiplos modelos (normalmente de tipos diferentes) e estatísticas simples (como a média) são usadas para combinar as previsões. O Voting Ensemble é um dos métodos Ensemble mais simples. Este método cria dois ou mais modelos separados a partir do dataset de treino. O Classificador Voting então utiliza a média das previsões de cada submodelo para fazer as previsões em novos conjuntos de dados. As previsões de cada submodelo podem receber pesos, através de parâmetros definidos manualmente ou através de heurística. Existem versões mais avançadas do Voting, em que o modelo pode aprender o melhor peso a ser atribuído aos submodelos. Isso é chamado Stacked Aggregation.