

## Rastgele Orman

Karar Ağacı (Decision Tree):

- Random Forest algoritması, temel olarak çok sayıda karar ağacının bir araya gelmesinden oluşur. Bu nedenle, öncelikle karar ağaçlarının nasıl çalıştığını anlamak önemlidir.
- Karar ağaçları, bir veri kümesindeki özellikleri (features) ve hedef değişkeni (target variable) arasındaki ilişkiyi modellemek için kullanılır. Her iç düğüm, bir özellikten bir eşik değeriyle bir karar noktası oluşturur ve her yaprak düğümü bir hedef değişkeni değeriyle sonuçlanır.
- Karar ağaçları, veri setini rekürsif olarak bölerek ve en iyi bölünme kriterlerini (genellikle bilgi kazancı veya jini indeksi kullanılarak) kullanarak veri kümesini homojen alt gruplara ayırarak çalışır.

Rastgele Orman (Random Forest):

- Random Forest, birden çok karar ağacının bir araya gelerek güçlü bir sınıflandırma veya regresyon modeli oluşturduğu bir topluluk öğrenme (ensemble learning) tekniğidir.
- Her karar ağacı, rastgele olarak özelliklerin bir alt kümesini ve örneklerin bir alt kümesini kullanarak eğitilir. Bu, her bir ağacın birbirinden farklı bir özellik alt kümesine odaklanmasını ve farklı veri noktalarına dayanarak farklı öğrenmeler yapmasını sağlar.
- Karar ağaçları genellikle özel durumlar ve aşırı uyum (overfitting) eğilimindedir. Ancak, Random Forest, birden çok karar ağacının tahminlerini bir araya getirerek bu eğilimleri azaltır ve daha genel geçer modeller oluşturur.
- Random Forest, her ağacın tahminlerinin ortalamasını (regresyon için) veya çoğunluk oylamasını (sınıflandırma için) kullanarak son tahmini yapar.

Bagging (Bootstrap Aggregating):

- Random Forest algoritması, bagging tekniğini kullanır. Bagging, birden çok modelin tahminlerini birleştirerek daha güçlü bir tahmin sağlayan bir topluluk öğrenme tekniğidir.
- Her karar ağacı, bootstrap örnekleme adı verilen rastgele örnekleme yöntemiyle eğitim verilerinden rastgele bir alt küme seçerek eğitilir.
- Son tahmin, tüm ağaçların tahminlerinin ortalaması veya çoğunluk oylamasıyla yapılır.

Random Forest, esnek, güçlü ve genel geçer bir sınıflandırma ve regresyon algoritmasıdır. Aşırı uydurmayı azaltır, rastgelelik ekleyerek çeşitlilik sağlar ve geniş veri kümeleri üzerinde yüksek performans gösterir.

## Uygulama

Pima Indian Diabetes veri setini kullanılmıştır.

Doğruluk: 0.7207792207792207

