

Random Forest

Rastgele Orman algoritması, Makine Öğreniminde güçlü bir ağaç öğrenme tekniğidir. Eğitim aşamasında bir dizi Karar Ağacı oluşturarak çalışır. Her ağaç, her bölümdeki özelliklerin rastgele bir alt kümesini ölçmek için veri kümesinin rastgele bir alt kümesi kullanılarak oluşturulur. Bu rastgelelik, tek tek ağaçlar arasında değişkenlik sağlayarak aşırı uyum riskini azaltır ve genel tahmin performansını artırır. Tahminde, algoritma, oylama (sınıflandırma görevleri için) veya ortalama (regresyon görevleri için) yoluyla tüm ağaçların sonuçlarını bir araya getirir. İlgörleriyle birden fazla ağaç tarafından desteklenen bu işbirliğine dayalı karar verme süreci, istikrarlı ve kesin sonuçlar için bir örnek sağlar. . Rastgele ormanlar, karmaşık verileri işleme, aşırı uyumu azaltma ve farklı ortamlarda güvenilir tahminler sağlama yetenekleriyle bilinen sınıflandırma ve regresyon fonksiyonları için yaygın olarak kullanılır.

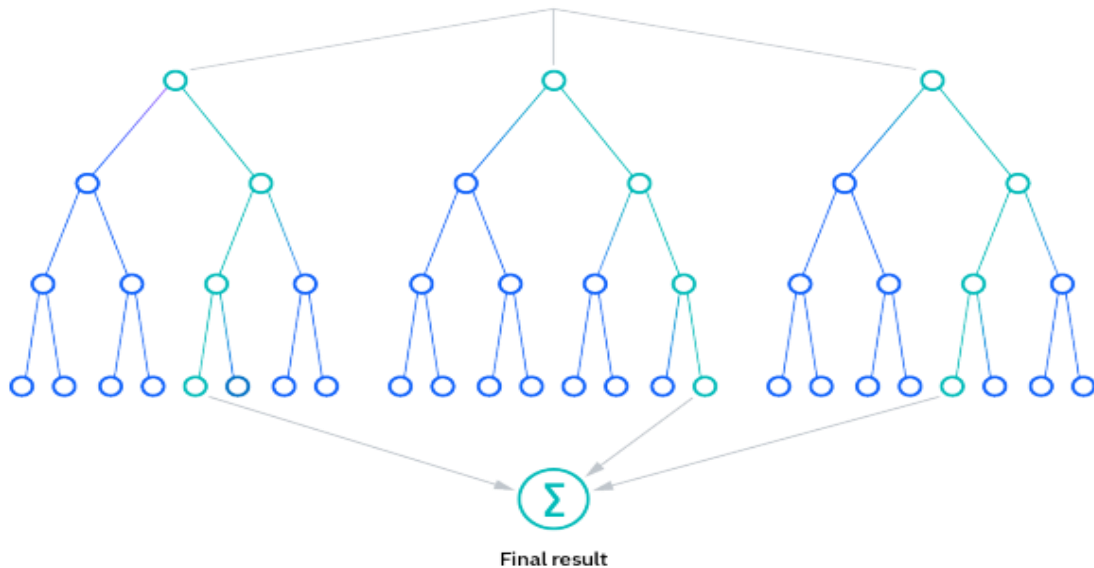
Nasıl Çalışır ?

Adım 1 : Algoritma, sağlanan veri kümesinden rastgele örnekler seçer.

Adım 2: Algoritma, seçilen her örnek için bir karar ağacı oluşturur. Ardından oluşturulan her karar ağacından bir tahmin sonucu alır.

Adım 3: Daha sonra tahmin edilen her sonuç için seçim yapılacaktır. Bir sınıflandırma probleminde regresyon problemi için ortalama kullanacaktır .

Adım 4 : Son olarak, algoritma, son tahmin olarak en çok oylanan tahmin sonucunu seçecektir.

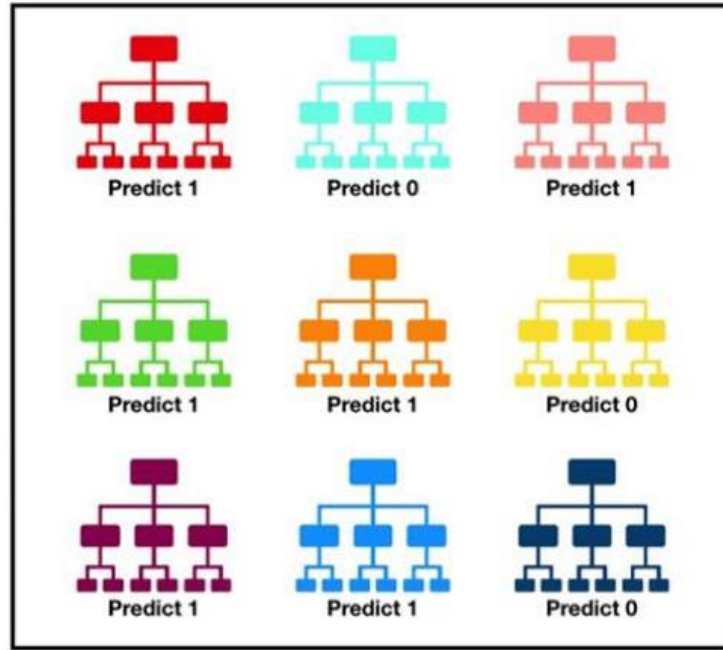


Bagging ve Boosting Nedir?

Bagging en çok oyu alan tahmin), birden fazla haftalık modellerin eğitim verilerinin farklı alt kümeleri üzerinde eğitildiği bir topluluk öğrenme modelidir. Her bir alt küme yer değiştirme ile örneklenir ve regresyon problemi için hafta modellerinin tahminlerinin ortalaması alınarak, sınıflandırma problemi için ise çoğunluk oyu dikkate alınarak tahmin yapılır.

Boosting, birden fazla tabanlı modeli sırayla eğitir. Bu yöntemde her model, önceki modellerin yaptığı hataları düzeltmeye çalışır. Her model, veri kümesinin değiştirilmiş bir versiyonu üzerinde eğitilir; önceki modeller tarafından yanlış sınıflandırılan örnekler daha fazla ağırlık verilir. Nihai tahmin ağırlıklı oylamayla yapılır

Özetle; rastgele orman adından da anlaşılacağı üzere bir topluluk olarak çalışan çok sayıda bireysel karar ağacından oluşur. **Ormandaki her bir ağaç bir sınıf tahmini verir ve en çok oyu alan sınıf, modelin öngörüsü haline gelir.**



Tally: Six 1s and Three 0s

6 tane 1 tahmin edilmiştir.

Random Forest Algoritmasının Avantajları:

- Birçok veri seti için en iyi doğruluk oranı ile tahmin de bulunur.
- Büyük veri tabanlarında çok etkili çalışır
- Çok fazla giriş değişkenini hane edebilir
- Kayıp verilerle mücadele konusunda etkilidir.