

Ağaç yapısı yapraklar ve karar düğümlerinden oluşur. Kökten başlayarak her karar düğümü girdiye bir bölme sınaması uygular ve sonuca göre dallardan biri alınır. Bir yaprağa ulaşıldığında arama durur. Burada örneğin hangi sınıfa ait olduğuna karar verilir.

Karar ağacı kurallardan oluşan bir kural tabanına dönüştürülebilir ve kuralların yorumlanmasını kolaylaştırır. Kara ağacı öğrenmesi parametrik değildir, böylece ağaç ihtiyaç duyulduğunda büyür (Alpaydın, 2019). Bu gibi durumlar karar ağaçlarının avantajları arasında yer alır.

Rastgele Orman

Rastgele ormanlar, birçok karar ağacının bir araya gelerek oluşturduğu bir ensemble (topluluk) yöntemidir. Her bir karar ağacı, farklı bir alt veri kümesi ve özellik kümesiyle eğitilir. Böylece, genellikle daha güçlü, daha kararlı ve daha az aşırı öğrenmeye eğilimli modeller üretilir.

Rastgele orman algoritmasına göre ilk olarak eğitim verisinden örnekler alınarak farklı alt kümeler oluşturulur. Her bir alt küme için bir karar ağacı oluşturulur. Bu süreçte her düğüm için rastgele bir özellik alt kümesi kullanılır. Bu durum aşırı uyum riskini azaltır. Sonraki aşamada tahmin edilen her bir hedef için oylar hesaplanır ve en yüksek oyu alan tahmin seçilir.

Rastgele orman algoritması hem regresyon hem de sınıflandırma problemlerinde kullanılabilir. Bu algoritma aşırı uyum problemini önlemektedir. Ayrıca mevcut özellikler arasından en önemli özelliği tanımlamak için de kullanılabilir (Çebi, 2020). Bu özellikler rastgele orman algoritmasının avantajları olarak gösterilir.

K-En Yakın Komşular (KNN)

Makine öğrenmesinde, sınıflandırma ve regresyon problemlerini çözmek için kullanılan basit ve etkili bir algoritmadır. Temel yaklaşım, bir veri noktasını etiketlemek veya tahmin etmek için, ona en yakın olan k komşusunun etiketlerini veya değerlerini kullanmaktır. En yakın noktalar bulunurken öklit, manhattan ve minkowski uzaklık değerlerine bakılır.

KNN algoritması uygulanması kolay bir algoritmaya sahip olması, sınıflandırma ve regresyon için kullanılabilmesi avantajları olarak görülmektedir. Ancak, bağımsız değişkenlerinin artmasıyla birlikte yavaşlaması, yüksek hafıza ihtiyacı, örnek sayısının artmasına paralel olarak tahmin süresinin artması dezavantajları olarak karşımıza çıkmaktadır (Atıcı 2022).

Destek Vektör Makinesi

Makine kavramsal olarak şu fikri uygular: giriş vektörleri, çok yüksek boyutlu bir özellik uzayına doğrusal olmayan bir şekilde eşlenir. Bu özellik uzayında doğrusal bir karar yüzeyi oluşturulur. Karar yüzeyinin özellikleri, makinenin yüksek genelleme yeteneği kazanmasını sağlar (Cortes, 1995). Destek vektör makinesi, veri noktaları arasındaki en büyük marjı (mesafe) maksimize etmeye çalışarak noktaları birbirinden en iyi şekilde ayıran, bir karar sınırı belirler. Veri doğrusal olarak ayrılabilirse bir hiper düzlem bulur, eğer veri doğrusal ayrılmıyorsa veriyi çok boyutlu özellik uzayına dönüştürerek doğrusal ayrıştırılabilir hale getirir.

Naive Bayes

Naive Bayes sınıflandırıcı, olasılığa dayalı bir makine öğrenimi sınıflandırma algoritmasıdır ve Bayes teoremini esas alır. Sınıflandırıcı bir eleman için her durumun olasılığını hesaplar ve olasılık değeri en yüksek olana göre sınıflandırır (Hatipoğlu 2023). Test kümesindeki bir veri eğitim kümesinde gözlemlenemiyorsa tahmin yapmaz. Bayes sınıflandırıcı, her özniteliğin birbirinden koşulsal bağımsız olduğu ve öğrenilmek istenen kavramın tüm bu özniteliklere koşulsal bağlı olduğu bir Bayes ağı olarak da düşünülebilir (Anonim, 2023).