

## Sınıflandırma Algoritması nedir ?

Sınıflandırma algoritması, makine öğreniminde kullanılan ve veri noktalarını belirli sınıflara ayırmayı amaçlayan bir tekniktir. Bu tür algoritmalar, genellikle denetimli öğrenme (supervised learning) yöntemlerine dayanır, yani model, önceden etiketlenmiş eğitim verileriyle eğitilir. Sınıflandırma, bir nesnenin belirli bir kategoriye veya sınıfa ait olup olmadığını belirlemeyi amaçlar.

Sınıflandırma algoritmalarının temel işleyişi şu şekildedir:

1. **Eğitim Aşaması:** Algoritma, etiketlenmiş veri seti üzerinde eğitilir. Bu veri seti, özellikler (features) ve bu özelliklere karşılık gelen etiketler (labels) içerir.
2. **Tahmin Aşaması:** Eğitilmiş model, yeni veri noktaları üzerinde kullanılarak bu noktaların hangi sınıfa ait olduğunu tahmin eder.

Sınıflandırma algoritmaları:

Diğer sınıflandırma tekniklerinden bazıları aşağıda verilmiştir:

- K-En Yakın Komşu Algoritması (K-Nearest Neighbour Algorithm)
- Lojistik Regresyon (Logistic Regression)
- Destek Vektör Makineleri (Support Vector Machines)
- Karar Ağaçları (Decision Tree)
- Rasgele Orman Kümeleri (Random Forests)
- Yapay Sinir Ağları (Artificial Neural Networks)
- Naive Bayes

Aşağıda bazı popüler sınıflandırma algoritmalarını ve temel özelliklerini bulabilirsiniz:

### 1. **Logistic Regression (Lojistik Regresyon):**

- o **Nasıl Çalışır:** İkili sınıflandırma problemleri için kullanılır. Verilerin belirli bir sınıfa ait olma olasılığını tahmin etmek için sigmoid fonksiyonunu kullanır.
- o **Kullanım Alanları:** Hastalık teşhisi, kredi riski analizi, pazarlama.

### 2. **Decision Trees (Karar Ağaçları):**

- o **Nasıl Çalışır:** Hem regresyon hem de sınıflandırma problemlerini çözmek için kullanılırlar. Karar ağacı, her yaprak düğümün bir sınıf etiketine karşılık geldiği ve özniteliklerin ağacın iç düğümünde temsil edildiği sorunu çözmek için ağaç temsili kullanır.
- o **Kullanım Alanları:** Müşteri segmentasyonu, kredi değerlendirme, biyoinformatik.

### 3. **Random Forests (Rastgele Ormanlar):**

- o **Nasıl Çalışır:** Birden fazla karar ağacının (ensemble) birleşiminden oluşur. Her ağaç, rastgele seçilen özellikler ve veri noktaları ile eğitilir ve son karar, ağaçların çoğunluk oyu ile belirlenir.
- o **Kullanım Alanları:** Hata ayıklama, finansal tahminler, sağlık teşhisi.

### 4. **Support Vector Machines (Destek Vektör Makineleri):**

- o **Nasıl Çalışır:** Veri noktalarını sınıflar arasında maksimum ayırım sağlayacak şekilde ayıran en iyi hiper düzlemi bulur. Lineer ve lineer olmayan sınıflandırma için kullanılabilir.
- o **Kullanım Alanları:** Görüntü tanıma, biyoinformatik, metin sınıflandırma.

### 5. **K-Nearest Neighbors (K-En Yakın Komşu):**

- o **Nasıl Çalışır:** Bu algoritma, yeni bir veri noktasının hangi sınıfa ait olduğunu belirlemek için, eğitim veri setindeki en yakın komşularını

kullanır. **Kullanım Alanları:** Kalp hastalığı tahmini, müşteri segmentasyonu, sahtekarlık tespiti.

#### 6. Naive Bayes:

- **Nasıl Çalışır:** Bayes teoremine dayanır ve özelliklerin birbirinden bağımsız olduğunu varsayar (naive assumption). Özelliklerin olasılıklarını kullanarak sınıf tahmini yapar.
- **Kullanım Alanları:** E-posta spam filtresi, belge sınıflandırma, duygu analizi.

#### 7. Neural Networks (Sinir Ağları):

- **Nasıl Çalışır:** Beynin çalışma şeklini taklit eden katmanlı yapıya sahip modellerdir. Özelliklerden yüksek seviyeli temsiller (representations) öğrenir ve bu temsiller üzerinden sınıflandırma yapar.
- **Kullanım Alanları:** Görüntü tanıma, ses tanıma, doğal dil işleme.

Sınıflandırma algoritmaları, geniş bir uygulama yelpazesine sahiptir ve günlük hayatta birçok alanda kullanılır. Hangi algoritmanın kullanılacağı, veri setinin yapısına, boyutuna ve problem tanımına bağlı olarak değişir. Ayrıca, modellerin performansını değerlendirmek ve iyileştirmek için çapraz doğrulama (cross-validation), ROC eğrileri ve F1 skoru gibi çeşitli değerlendirme metrikleri kullanılır.