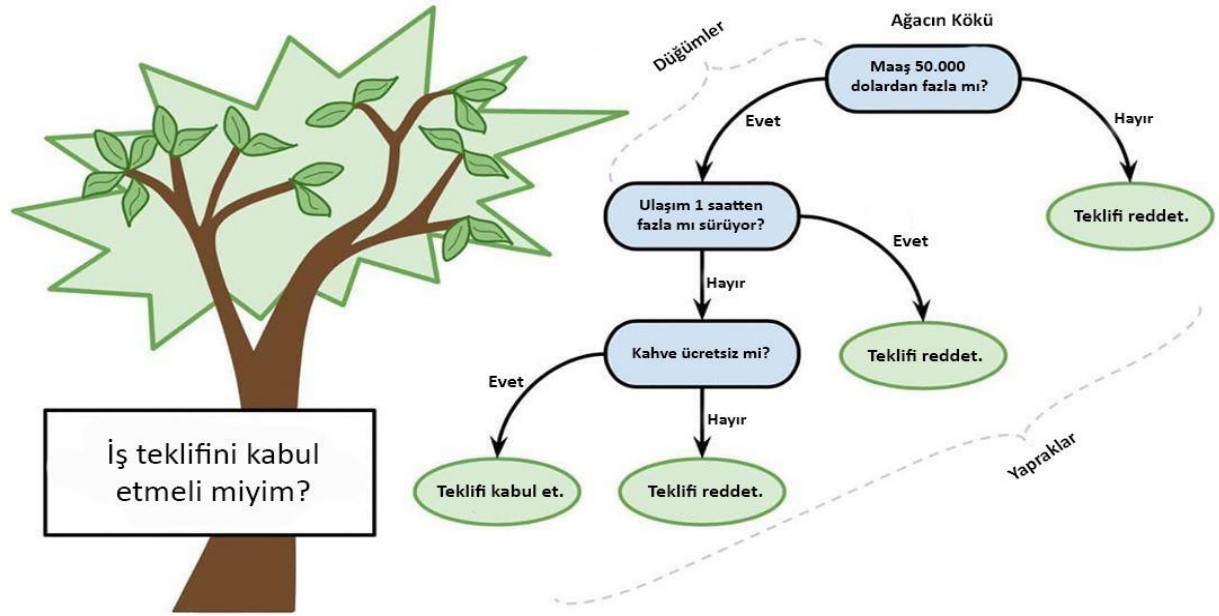


## Karar Ağaçları (Decision Trees)



**Karar ağaçları**, veri noktalarını özelliklerine dayalı olarak dallanma yapısıyla sınıflandıran veya tahmin eden denetimli öğrenme algoritmalarıdır. Hem sınıflandırma hem de regresyon problemlerinde kullanılabilirler. Karar ağaçları, her düğümde bir karar verir ve her dalda veri noktalarını bu karara göre böler.

### Temel Kavramlar

1. **Kök Düğüm (Root Node):** Karar ağacının başladığı noktadır ve tüm veri setini temsil eder.
2. **Dahili Düğümler (Internal Nodes):** Veri setini alt gruplara ayıran karar noktalarıdır.
3. **Yaprak Düğümler (Leaf Nodes):** Nihai sınıflandırmaları veya tahminleri temsil eder.
4. **Dallanma (Splitting):** Veri setini alt gruplara ayırma işlemi.
5. **Budama (Pruning):** Aşırı uyumu önlemek için gereksiz dalların kaldırılması işlemi.

### Karar Ağaçları Nasıl Çalışır?

#### 1. Bölünme Kriteri Seçimi:

- Karar ağaçları, veri setini en iyi şekilde nasıl böleceğini belirlemek için belirli kriterler kullanır. Bu kriterler, Gini impurity, bilgi kazancı (information gain), veya varyans azalması gibi ölçütler olabilir.

#### 2. Bölünme İşlemi:

- Her düğümde, veri seti belirli bir özelliğe göre bölünür. Bu bölünme, belirlenen kriterlere göre yapılır.

#### 3. Ağaç Oluşturma:

- Veri seti, bölünme işlemi ile sürekli olarak alt gruplara ayrılarak ağaç oluşturulur. Bu işlem, tüm veriler yaprak düğümlere ayrılana veya belirli bir durdurma kriterine ulaşılan kadar devam eder.

#### 4. Budama:

- Aşırı uyumu önlemek için ağaç budanır. Bu, fazla dalların kaldırılması anlamına gelir ve modelin genelleştirme yeteneğini artırır.

### Örnek 1: Sınıflandırma Problemi

Bir veri seti düşünelim: Bir mağazada müşteri verileri var ve müşterilerin bir ürünü satın alıp almayacağını tahmin etmek istiyoruz. Veri setinde yaş, gelir seviyesi, cinsiyet gibi özellikler mevcut.

#### 1. Adım 1 - Kök Düzüm:

- Kök düğüm, tüm veri setini temsil eder. Örneğin, yaşa göre ilk bölünmeyi yapabiliriz.

#### 2. Adım 2 - Dahili Düzümler:

- Yaşa göre veri seti iki gruba ayrılır: 30 yaşın altındakiler ve 30 yaşın üstündekiler.
- 30 yaşın altındaki grup, gelir seviyesine göre yeniden bölünür: düşük gelir ve yüksek gelir.

#### 3. Adım 3 - Yaprak Düzümler:

- Her grup için nihai kararlar verilir. Örneğin, 30 yaşın altında ve düşük gelirli müşterilerin ürünü satın alma olasılığı düşük, 30 yaşın üstünde ve yüksek gelirli müşterilerin ürünü satın alma olasılığı yüksek olabilir.

### Bölünme Kriteri Örneği:

- **Bilgi Kazancı (Information Gain):** Veri setindeki belirsizliği azaltma miktarına dayanır. Daha fazla bilgi kazancı sağlayan bölünme, daha iyi bir bölünmedir.
- **Gini Impurity:** Belirli bir düğümdeki yanlış sınıflandırma olasılığını ölçer. Daha düşük Gini impurity, daha iyi bir bölünmeyi gösterir.

### Örnek 2: Regresyon Problemi

Bir veri seti düşünelim: Bir gayrimenkul firması, ev fiyatlarını tahmin etmek istiyor. Veri setinde evin metrekaresi, oda sayısı, konumu gibi özellikler mevcut.

#### 1. Adım 1 - Kök Düzüm:

- Kök düğüm, tüm veri setini temsil eder. Örneğin, evin metrekaresine göre ilk bölünmeyi yapabiliriz.

#### 2. Adım 2 - Dahili Düzümler:

- Metrekareye göre veri seti iki gruba ayrılır: 100 metrekarenin altındaki evler ve 100 metrekarenin üstündeki evler.

- 100 metrekarenin altındaki grup, oda sayısına göre yeniden bölünür: 2 oda ve daha az, 3 oda ve daha fazla.

### 3. Adım 3 - Yaprak Düğümler:

- Her grup için nihai tahminler yapılır. Örneğin, 100 metrekarenin altında ve 2 odalı evlerin ortalama fiyatı X lira, 100 metrekarenin üstünde ve 3 odalı evlerin ortalama fiyatı Y lira olabilir.

### Bölünme Kriteri Örneği:

- **Varyans Azalması (Variance Reduction):** Bölünmenin hedef değişkenin varyansını ne kadar azalttığını ölçer. Daha büyük varyans azalması, daha iyi bir bölünmeyi gösterir.

### Karar Ağaçlarının Avantajları ve Dezavantajları

#### Avantajları:

- **Görselleştirme ve Yorumlanabilirlik:** Karar ağaçları, sonuçlarını görsel olarak sunabildiği için kullanıcılar tarafından kolayca anlaşılabilir.
- **Özellik Seçimi Gereksinimi Azdır:** Önemli özellikleri otomatik olarak seçer.
- **Kategorik ve Sayısal Verilerle Çalışma Yeteneği:** Hem kategorik hem de sayısal verilerle etkili şekilde çalışır.

#### Dezavantajları:

- **Aşırı Uyuma Eğilimi:** Karar ağaçları, veri setine fazla uyum sağlama riski taşır, bu da genelleştirme yeteneğini azaltır.
- **Dengesiz Veri Setlerinde Performans Düşüşü:** Dengesiz sınıf dağılımlarında kötü performans gösterebilir.
- **Karmaşıklık:** Çok derin ağaçlar, büyük ve karmaşık modeller oluşturabilir.

### Örnek Bir Karar Ağacı

**Veri Seti:** Bir bahçedeki bitkilerin türlerini tahmin etmek için bitki boyu ve yaprak genişliği gibi özelliklere sahip bir veri seti.

1. **Kök Düğüm:** İlk bölünme noktası olarak bitki boyunu seçiyoruz. Örneğin, 100 cm.
2. **Dahili Düğümler:**
  - 100 cm'in altında olan bitkileri, yaprak genişliğine göre yeniden bölüyoruz: 5 cm ve daha az, 5 cm'den fazla.
3. **Yaprak Düğümler:** Her bir alt grup için bitki türünü belirliyoruz. Örneğin, 100 cm'in altında ve yaprak genişliği 5 cm'den az olan bitkiler A türü, 100 cm'in üstünde ve yaprak genişliği 5 cm'den fazla olan bitkiler B türü olabilir.

Bu süreç, tüm veri seti yaprak düğümlere ayrılana kadar devam eder. Nihayetinde, her bir yaprak düğüm bir sınıfı veya tahmini temsil eder ve karar ağacı tamamlanır.

Bu örnekler ve açıklamalarla, karar ağaçlarının nasıl çalıştığını ve nasıl kullanıldığını daha iyi anlayabilirsiniz. Karar ağaçları, veri analizi ve tahmin problemlerinde güçlü ve esnek bir araçtır, ancak doğru ayar ve budama teknikleriyle kullanıldığında en iyi sonuçları verir.