

CART Denum CAUSMASI

is weekend	weather	parents	money	Decision
w1	Sunny	Y	Rich	Cinema
w2	Sunny	N	Rich	Tennis
w3	Windy	Y	Rich	Cinema
w4	Rainy	Y	Poor	Cinema
w5	Rainy	N	Rich	stay I
w6	Rainy	Y	Poor	Cinema
w7	Windy	N	Poor	Cinema
w8	Windy	N	Rich	Shopping
w9	Windy	Y	Rich	Cinema
w10	Sunny	N	Rich	Tennis
	3 kez	2 kez	2 kez	target
	feature			target

money (poor) = 3 cinema, 0 T, 0 St, 0 shop

$$Gini_{poor} = 1 - \left[\left(\frac{3}{3} \right)^2 + \left(\frac{0}{3} \right)^2 + \left(\frac{0}{3} \right)^2 + \left(\frac{0}{3} \right)^2 \right] = 0$$

money (Rich) = 3 cinema, 2 tennis, 1 stay, 1 shop

$$Gini_{rich} = 1 - \left[\left(\frac{3}{7} \right)^2 + \left(\frac{2}{7} \right)^2 + \left(\frac{1}{7} \right)^2 + \left(\frac{1}{7} \right)^2 \right] = 0.68$$

$$Gini(money) = \text{Ağırlıklı Ortalama} = 0 \cdot \frac{3}{10} + 0.68 \cdot \frac{7}{10} = 0.48 \rightarrow \text{money feature gini değeri}$$

Parents (Yes) = 5 cinema, 0 T, 0 St, 0 shop

$$Gini_{yes} = 0$$

Parents (No) = 2 tennis, 1 cinema, 1 shop, 1 stay

$$Gini_{no} = 1 - \left[\left(\frac{2}{5} \right)^2 + \left(\frac{1}{5} \right)^2 + \left(\frac{1}{5} \right)^2 + \left(\frac{1}{5} \right)^2 \right] = 0.72$$

$$Gini(Parents) = 0 \cdot \frac{5}{10} + 0.72 \cdot \frac{5}{10} = 0.36$$

Weather (Sunny) = 1 cinema 2 tennis

$$Gini_{Sunny} = 1 - \left[\left(\frac{1}{3} \right)^2 + \left(\frac{2}{3} \right)^2 \right] = 0.44$$

Weather (windy) = 3 cinema 1 shopping

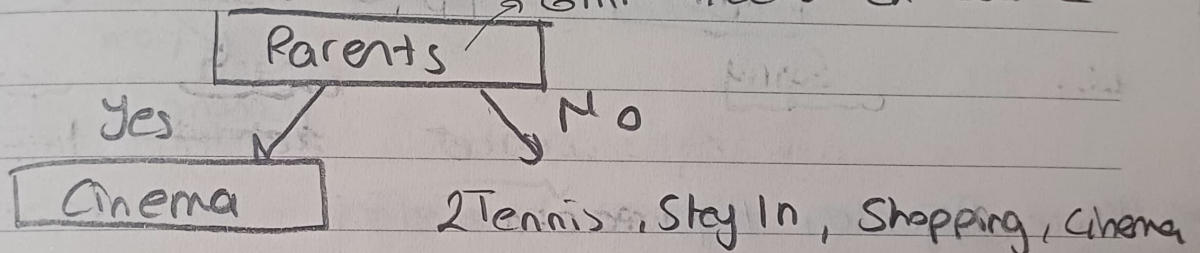
$$Gini_{windy} = 1 - \left[\left(\frac{3}{4} \right)^2 + \left(\frac{1}{4} \right)^2 \right] = 0.37$$

Weather (Rainy) = 2 cinema 1 stay in

$$Gini_{rainy} = 0.44$$

$$Gini(weather) = 0.44 \cdot \frac{3}{10} + 0.37 \cdot \frac{4}{10} + 0.44 \cdot \frac{3}{10} = 0.41$$

= 0.41 → Gini indeksi en düşük



Sadece No ile verisetinde ayrı işlemleri tercih yap.

Weekend	Weather	Parents	Money	Decision
w1	Sunny	No	Rich	Tennis
w5	Rainy	N	Rich	Stay In
w7	windy	N	Poor	Cinema
w8	windy	N	Rich	Shopping
w10	Sunny	N	Rich	Tennis
	0.2		0.5	

Money (Rich) = 2 tennis 1 stay 1 shop

$$Gini_{Rich} = 1 - \left[\left(\frac{2}{4} \right)^2 + \left(\frac{1}{4} \right)^2 + \left(\frac{1}{4} \right)^2 \right] = 0.625$$

Money (Poor) = 1 Cinema = 0

$$Gini(money) = 0.625 \cdot \frac{4}{5} + 0 \cdot \frac{1}{5} = 0.5$$

$$\text{weather (Sunny)} = 2 \text{ tennis}$$

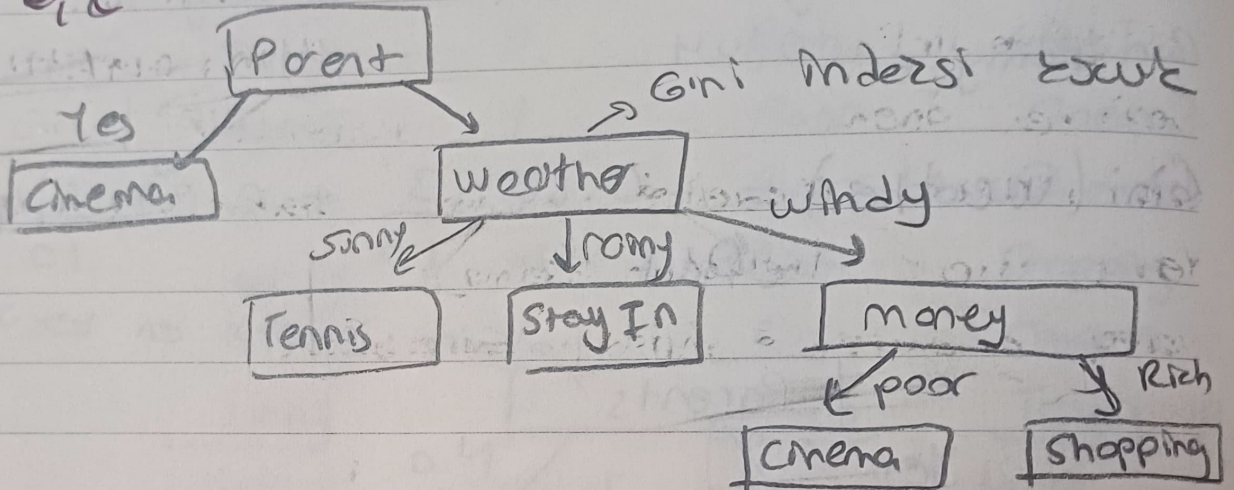
$$\text{Gini sunny} = 0$$

$$\text{weather (Rainy)} = 1 \text{ stay in } \text{Gini rainy} = 0$$

$$\text{weather (windy)} = 1 \text{ chess } 1 \text{ shopping}$$

$$\text{Gini windy} = 1 - \left(\left(\frac{1}{2} \right)^2 + \left(\frac{1}{2} \right)^2 \right) = 0,5$$

$$\text{Gini (weather)} = 0 \cdot \frac{2}{5} + 0 \cdot \frac{2}{5} + 0,5 \cdot \frac{2}{5} = 0,2$$



Karar Ağaçları Hiperparametreleri

- 1) max - depth = max derinlik
 - 2) min - samples - split (min bölme boyutu) : Bir düğümün bölünmesi için gereken min veri noktası sayısını belirler. Ağacın bölünme yapmadan önce ne kadar veri toplamasını gerektirdiğini belirler.
 - 3) min - samples - leaf (min yaprak boyutu) : Bir düğümde bulunması gereken min veri noktasını belirler.
 - 4) max - features (max özellikler) : Bir düğümde değerlendirecek max özellik sayısını belirler.
 - 5) criterion (Bölünme kriteri) : Gini ve Entropy'dir.
- Bu hiperparametrelerin ayarlanması için kross validation ve grid search yöntemleri kullanılır.

Decision Tree Uygulama Alanları

- Sınıflandırma Problemleri (Kategorik veridible)
- Regresyon Problemleri (Nümerik veridible) ↑ kategorik gibi düşün
- Pazarlama ve Satış Analizi
- Risk Değerlendirmesi
- Söplü ve T, P
- Endüstriyel Üretim ve Kalite Kontrol

Avantajları

Anlaşılabilirlik

Derinden öğrenme

Gözle görülür modellerin uygulanması

Veri ön işleme ihtiyacının azalması

Kategorik verilerle çalışmada kolaylık

Dezavantajları

Aşırı uyum (overfitting)

Yaygınlık

Dengesiz veri kümeleri