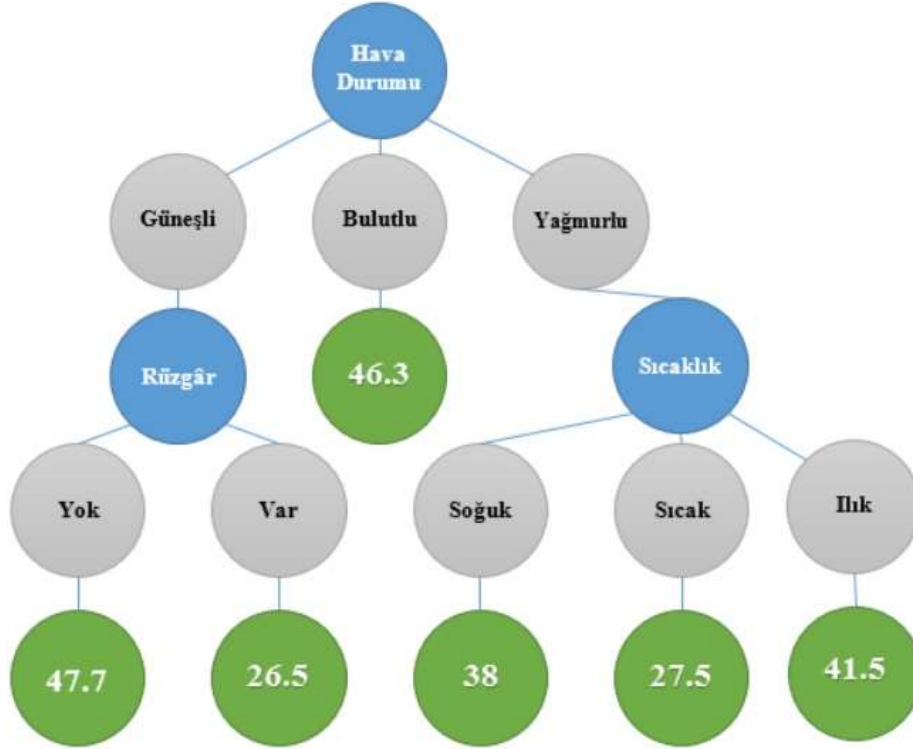


05.01. Decision Tree – Regression (Regresyon)

Karar ağaçları hem regresyon hem de sınıflama modelleri üzerine inşa edilebilen ağaç yapısı formundadır. Regresyon sayısal hedef verisi üzerinde kullanılırken sınıflama ise kategorik veriler (Örneğin evet/hayır) üzerinde kullanılır. Karar ağaçları, özellik ve hedefe göre karar düğümleri (decision nodes) ve yaprak düğümlerinden (leaf nodes) oluşur. Aşağıdaki verileri ve karar ağacını incelerken Decision Tree – Classification veri ve ağaç çizimi ile karşılaştırmanızı öneririm.

Özellikler				Hedef
Hava Durumu	Sıcaklık	Nem	Rüzgar	Futbol Oyna
Yağmurlu	Sıcak	Yüksek	Yok	25
Yağmurlu	Sıcak	Yüksek	Var	30
Bulutlu	Sıcak	Yüksek	Yok	46
Güneşli	Ilık	Yüksek	Yok	45
Güneşli	Soğuk	Normal	Yok	52
Güneşli	Soğuk	Normal	Var	23
Bulutlu	Soğuk	Normal	Var	43
Yağmurlu	Ilık	Yüksek	Yok	35
Yağmurlu	Soğuk	Normal	Yok	38
Güneşli	Ilık	Normal	Yok	46
Yağmurlu	Ilık	Normal	Yok	48
Bulutlu	Ilık	Yüksek	Var	52
Bulutlu	Sıcak	Normal	Yok	44
Güneşli	Ilık	Yüksek	Var	30



Hedef özelliği kaç dakika futbol oynayacağımız hakkında bilgi veriyor. Karar ağacında ise verilen özelliğe göre kaç dakika futbol oynayacağımızın sonucunu görüyoruz. Bu sonuç elimizdeki eğitim verisine göre yaklaşık sonuç verdiğini görmüşünüzdür. Şimdi bu sonucu nasıl elde edeceğimize bakalım.

Decision Tree Algoritması: ID3

Bu bölümü daha önceki karar ağacı – sınıflama yazımızdaki ID3 algoritması üzerinden anlatacağız. Regresyon sebebiyle information gain yerine standart sapma kullanacağız. Öncelikle hedef için standart sapmayı hesaplayalım.

Hedef
Futbol Oyna
25
30
46
45
52
23
43
35
38
46
48
52
44
30

$$S = \sqrt{\frac{\sum (x - \mu)^2}{n}} = 9.32$$

Ayrıca, özellik ve hedef ikilileri için standart sapmaları hesaplayalım. İki parametrelili standart sapma hesaplama da aşağıdaki formülasyonlar kullanılır.

$$S(T, X) = \sum_{c \in X} P(c) S(c)$$

$$SDR(T, X) = S(T) - S(T, X)$$

Örneğin hava durumu ve futbol oyna ikilisi için hesaplama yaparsak:

$$\begin{aligned} S(\text{Hava Durumu}, \text{Futbol Oyna}) &= P(\text{Güneşli}) * S(\text{Güneşli}) + P(\text{Bulutlu}) * S(\text{Bulutlu}) + P(\text{Yağmurlu}) * S(\text{Yağmurlu}) \\ &= 5/14 * 10.870 + 4/14 * 3.491 + 5/14 * 7.782 \\ &= 7.659 \end{aligned}$$

$$\begin{aligned} SDR(\text{Hava Durumu}, \text{Futbol Oyna}) &= S(\text{Güneşli}) - S(\text{Hava Durumu}, \text{Futbol Oyna}) \\ &= 1.662 \end{aligned}$$

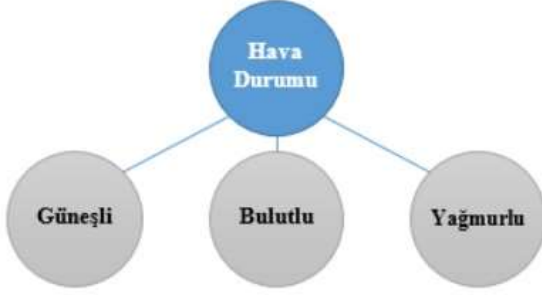
		Futbol Oyna			
		Standart Sapma	Toplam	P(c) * S(c)	
Hava Durumu	Güneşli	10.870	5	3.882	P(Güneşli) * S(Güneşli)
	Bulutlu	3.491	4	0.997	P(Bulutlu) * S(Bulutlu)
	Yağmurlu	7.782	5	2.779	P(Yağmurlu) * S(Yağmurlu)
			14	7.659	Toplam
			SDR	1.662	

		Futbol Oyna			
		Standart Sapma	Toplam	P(c) * S(c)	
Sıcaklık	Sıcak	8.955	4	2.558	
	Ilık	7.652	6	3.279	
	Soğuk	10.512	4	3.003	
			14	8.841	
			SDR	0.480	

		Futbol Oyna			
		Standart Sapma	Toplam	P(c) * S(c)	
Nem	Yüksek	9.363	7	4.682	
	Normal	8.734	7	4.367	
			14	9.049	
			SDR	0.272	

		Futbol Oyna		
		Standart Sapma	Toplam	P(c)*S(c)
Rüzgar	Var	10.442	5	3.729
	Yok	7.709	9	4.956
			14	8.685
			SDR	0.636

SDR değeri en büyük olan Hava Durumu özelliği kök olarak seçilir. Bu durumda veri tekrar ele alınır. Daha anlaşılır olması açısından veriyi Hava Durumu özelliği göre gruplandırıldı.



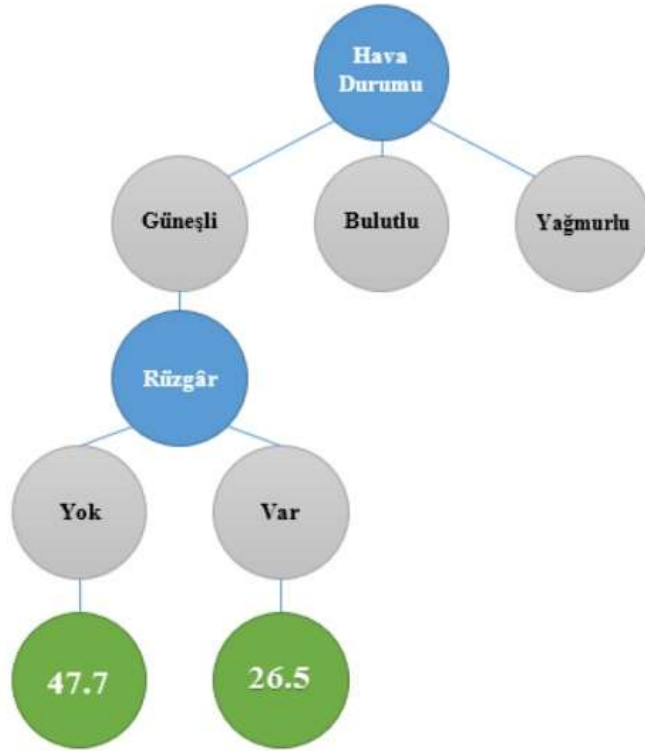
Özellikler				Hedef
Hava Durumu	Sıcaklık	Nem	Rüzgar	Futbol Oyna
Güneşli	Ilık	Yüksek	Yok	45
Güneşli	Soğuk	Normal	Yok	52
Güneşli	Soğuk	Normal	Var	23
Güneşli	Ilık	Normal	Yok	46
Güneşli	Ilık	Yüksek	Var	30
Bulutlu	Sıcak	Yüksek	Yok	46
Bulutlu	Soğuk	Normal	Var	43
Bulutlu	Ilık	Yüksek	Var	52
Bulutlu	Sıcak	Normal	Yok	44
Yağmurlu	Sıcak	Yüksek	Yok	25
Yağmurlu	Sıcak	Yüksek	Var	30
Yağmurlu	Ilık	Yüksek	Yok	35
Yağmurlu	Soğuk	Normal	Yok	38
Yağmurlu	Ilık	Normal	Yok	48

Bu durumda teker teker Güneşli, Bulutlu ve Yağmurlu verileri için SDR'ler tekrar hesaplanır. Güneş için SDR'leri hesaplayalım. Sıcaklık, Nem ve Rüzgar için SDR hesaplandığımızda:

Özellikler				Hedef
Hava Durumu	Sıcaklık	Nem	Rüzgar	Futbol Oyna
Güneşli	Ilık	Yüksek	Yok	45
Güneşli	Soğuk	Normal	Yok	52
Güneşli	Soğuk	Normal	Var	23
Güneşli	Ilık	Normal	Yok	46
Güneşli	Ilık	Yüksek	Var	30
		S.Sapma		10.870
		Futbol Oyna		
		Standart Sapma	Toplam	P(c)*S(c)
Rüzgar	Var	3.500	2	0.636
	Yok	3.091	9	2.529
			11	3.166
			SDR	7.705

		Futbol Oyna		
		Standart Sapma	Toplam	P(c) * S(c)
Sıcaklık	Ilık	7.318	3	4.391
	Soğuk	14.500	2	5.800
			5	10.191
			SDR	0.679
		Futbol Oyna		
		Standart Sapma	Toplam	P(c) * S(c)
Nem	Yüksek	7.500	2	3.000
	Normal	12.499	3	7.499
			5	10.499
			SDR	0.371

En yüksek SDR değeri 7.705 ile rüzgar için çıkmıştır. Bu durumda ağaca rüzgar eklenir. Bu duruma gelindiğinde bir bitirme kriterinin belirtilmesi gerekmektedir. Örneğin %5 verinin altına düşündüğünde bölme işlemini bırak. Rüzgar eklendikten belirlenen kriter sebebiyle bölme işlemini bitirilir. Bu noktada Rüzgar verisindeki var ve yok'lar için ortalama değer hesaplanır ve aşağıdaki karar ağacı elde edilir.



Öz yinelemeli şekilde tüm alt düğümler hesaplanır ve karar ağacı elde edilir.

[Decision_Tree_Reg.xlsx \(14,32 kb\)](#)

Makine Öğrenmesi

[« Önceki: 05. Regresyon](#)

[Sonraki: 07. Kümeleme »](#)



Search...

- 01. Makine Öğrenmesi
- 02. Supervised ve Unsupervised Learning
- 03. Model Başarısı Değerlendirme – Sınıflandırma
- 04. Sınıflandırma

- [05. Regresyon](#)
 - **05.01. Decision Tree – Regression (Regresyon)**
- [07. Kümeleme](#)
- [08. Association Rules \(Birliktelik Kuralları\)](#)