

Podana je učna množica 14 učnih primerov, ki podaja vremenske pogoje in število ur igranja tenisa v lokalnem teniškem klubu. Kateri atribut bo v korenu drevesa? Zgradi regresijsko odločitveno drevo, ki iz vremenskih (opisnih) podatkov napove pričakovano število ur igranja tenisa.

Vreme	Temperatura	Vlažnost	Veter	Ure igranja
deževno	vroče	visoka	ne	25
deževno	vroče	visoka	da	30
oblačno	vroče	visoka	ne	46
sončno	prijetno	visoka	ne	45
sončno	hladno	normalna	ne	52
sončno	hladno	normalna	da	23
oblačno	hladno	normalna	da	43
deževno	prijetno	visoka	ne	35
deževno	hladno	normalna	ne	38
sončno	prijetno	normalna	ne	46
deževno	prijetno	normalna	da	48
oblačno	prijetno	visoka	da	52
oblačno	vroče	normalna	ne	44
sončno	prijetno	visoka	da	30

UI = ure igranja

UI = [25, 30, 46, 45, 52, 23, 43, 35, 38, 46, 48, 52, 44, 30]

$$\mu = 39,79$$

$$MSE(UI) = \frac{1}{14} ((25-39,79)^2 + (30-39,79)^2 + (46-39,79)^2 + (45-39,79)^2 + (52-39,79)^2 + (23-39,79)^2 + (43-39,79)^2 + (35-39,79)^2 + (38-39,79)^2 + (46-39,79)^2 + (48-39,79)^2 + (52-39,79)^2 + (44-39,79)^2 + (30-39,79)^2)$$

$$MSE(UI) = 86,88$$

Za koren izberemo atribut, ki razdeli množico tako, da je rezidualna kvadratna napaka najmanjša.

Vreme

Atribut *vreme* ima tri vrednosti: deževno (d), sončno (s) in oblačno (o). Atribut *vreme* razdeli množico *UI* na naslednje 3 podmnožice.

Vreme = oblačno

$$UI_{vo} = [46, 43, 52, 44]$$

$$\mu_{vo} = 46,25$$

$$MSE(UI_{vo}) = \frac{1}{4} ((46-46,25)^2 + (43-46,25)^2 + (52-46,25)^2 + (44-46,25)^2)$$

$$MSE(UI_{vo}) = 12,19$$

$$p_{vo} = \frac{4}{14}$$

Vreme = deževno

$$UI_{vd} = [25, 30, 35, 38, 48]$$

$$\mu_{vd} = 35,20$$

$$MSE(UI_{vd}) = \frac{1}{5} ((25-35,20)^2 + (30-35,20)^2 + (35-35,20)^2 + (38-35,20)^2 + (48-35,20)^2)$$

$$MSE(UI_{vd}) = 60,56$$

$$p_{vd} = \frac{5}{14}$$

Vreme = sončno

$$UI_{vs} = [45, 52, 23, 46, 30]$$

$$\mu_{vs} = 39,20$$

$$MSE(UI_{vs}) = \frac{1}{5} ((45-39,20)^2 + (52-39,20)^2 + (23-39,20)^2 + (46-39,20)^2 + (30-39,20)^2)$$

$$MSE(UI_{vs}) = 118,16$$

$$p_{vs} = \frac{5}{14}$$

$$MSE_{res}(vreme) = p_{vo} MSE(UI_{vo}) + p_{vd} MSE(UI_{vd}) + p_{vs} MSE(UI_{vs})$$

$$MSE_{res}(vreme) = 67,31$$

Temperatura

Atribut *temperatura* ima tri vrednosti: vroče (v), prijetno (p) in hladno (h). Atribut *temperatura* razdeli množico *UI* na naslednje 3 podmnožice.

Temperatura = vroče

$$UI_{tv} = [25, 30, 46, 44]$$

$$\mu_{tv} = 36,25$$

$$MSE(UI_{tv}) = \frac{1}{4} ((25-36,25)^2 + (30-36,25)^2 + (46-36,25)^2 + (44-36,25)^2)$$

$$MSE(UI_{tv}) = 80,19$$

$$p_{tv} = \frac{4}{14}$$

Temperatura = prijetno

$$UI_{tp} = [45, 35, 46, 48, 52, 30]$$

$$\mu_{tp} = 42,67$$

$$MSE(UI_{tp}) = \frac{1}{6} ((45-42,67)^2 + (35-42,67)^2 + (46-42,67)^2 + (48-42,67)^2 + (52-42,67)^2 + (30-42,67)^2)$$

$$MSE(UI_{tp}) = 58,56$$

$$p_{tp} = \frac{6}{14}$$

Temperatura = hladno

$$UI_{th} = [52, 23, 43, 38]$$

$$\mu_{th} = 39,00$$

$$MSE(UI_{th}) = \frac{1}{4} ((52-39)^2 + (23-39)^2 + (43-39)^2 + (38-39)^2)$$

$$MSE(UI_{th}) = 110,50$$

$$p_{th} = \frac{4}{14}$$

$$MSE_{res}(temperatura) = p_{tv} MSE(UI_{tv}) + p_{tp} MSE(UI_{tp}) + p_{th} MSE(UI_{th})$$

$$MSE_{res}(temperatura) = 79,58$$

Vlažnost

Atribut *vlažnost* ima dve vrednosti: visoka (v) in normalno (n). Atribut *vlažnost* razdeli množico *UI* na nasledni 2 podmnožici.

Vlažnost = visoka

$$UI_{vv} = [25, 30, 46, 45, 35, 52, 30]$$

$$\mu_{vv} = 37,57$$

$$MSE(UI_{vv}) = 1/7 ((25-37,57)^2 + (30-37,57)^2 + (46-37,57)^2 + (45-37,57)^2 + (35-37,57)^2 + (52-37,57)^2 + (30-37,57)^2)$$

$$MSE(UI_{vv}) = 87,67$$

$$p_{vv} = 7/14$$

Vlažnost = normalna

$$UI_{vn} = [52, 23, 43, 38, 46, 48, 44]$$

$$\mu_{vn} = 42,00$$

$$MSE(UI_{vn}) = 1/7 ((52-42)^2 + (23-42)^2 + (43-42)^2 + (38-42)^2 + (46-42)^2 + (48-42)^2 + (44-42)^2)$$

$$MSE(UI_{vn}) = 76,69$$

$$p_{vn} = 7/14$$

$$MSE_{res}(vlažnost) = p_{vv} MSE(UI_{vv}) + p_{vn} MSE(UI_{tn})$$

$$MSE_{res}(vlažnost) = 79,58$$

Veter

Atribut *veter* ima dve vrednosti: da in ne. Atribut *veter* razdeli množico *UI* na nasledni 2 podmonožici.

Veter = da

$$UI_{vda} = [30, 23, 43, 48, 52, 30]$$

$$\mu_{vda} = 37,67$$

$$MSE(UI_{vda}) = \frac{1}{6} ((30-37,67)^2 + (23-37,67)^2 + (43-37,67)^2 + (48-37,67)^2 + (52-37,67)^2 + (30-37,67)^2)$$

$$MSE(UI_{vda}) = 112,22$$

$$p_{vda} = \frac{6}{14}$$

Veter = ne

$$UI_{vne} = [25, 46, 45, 52, 35, 38, 46, 44]$$

$$\mu_{vne} = 41,38$$

$$MSE(UI_{vne}) = \frac{1}{8} ((25-41,38)^2 + (46-41,38)^2 + (45-41,38)^2 + (52-41,38)^2 + (35-41,38)^2 + (38-41,38)^2 + (46-41,38)^2 + (44-41,38)^2)$$

$$MSE(UI_{vne}) = 61,98$$

$$p_{vne} = \frac{8}{14}$$

$$MSE_{res}(veter) = p_{vda} MSE(UI_{vda}) + p_{vne} MSE(UI_{vne})$$

$$MSE_{res}(veter) = 83,51$$

Atribut *vreme* ima najmanjši *mse*, zato ga zberemo, da gre v koren drevesa. Gradnjo drevesa nadaljujemo na podoben način.