

19360859053

HÜMEYRA ÇİMEN

1.04.2023

2023-2024 EĞİTİM ÖĞRETİM YILI

BURSA TEKNİK ÜNİVERSİTESİ MÜHENDİSLİK VE DOĞA BİLİMLER FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ VERİ MADENCİLİĞİ DERSİ ÖDEV-1

A)

$$\text{Gini} = 1 - p(C0/C0+C1) * 2 - p(C1/C0+C1) * 2 = 1 - (10/20)^2 - (10/20)^2 = 1 - 0,25 - 0,25 = 0,5$$

$$\text{Entropi} = - (10/20) \log_2(10/20) - (10/20) \log_2(10/20) = 1$$

B)

$$\text{Gini} = 1 - (1/1)^2 - (0/1)^2 = 0$$

$$\text{Entropi} = - (1/1) \log_2(1/1) - (0/1) \log_2(0/1) = 0$$

C)

$$\text{Gini}(M) = 1 - p(C0/M)^2 - p(C1/M)^2 = 1 - (6/10)^2 - (4/10)^2 = 0,48$$

$$\text{Gini}(F) = 1 - p(C0/F)^2 - p(C1/F)^2 = 1 - (4/10)^2 - (6/10)^2 = 0,48$$

$$\text{Gini} = [(M) / (M+F)] * \text{Gini}(M) + [(F) / (M+F)] * \text{Gini}(F) = (10/20) * 0,48 + (10/20) * 0,48 = 0,48$$

$$\text{Entropi}(M) = - (6/10) \log_2(6/10) - (4/10) \log_2(4/10) = 0,971$$

$$\text{Entropi}(F) = - (4/10) \log_2(4/10) - (6/10) \log_2(6/10) = 0,971$$

$$\text{Entropi} = (10/20) * 0,971 + (10/20) * 0,971 = 0,971$$

D)

$$\text{Gini}(\text{Family}) = 1 - (1/4)^2 - (3/4)^2 = 0,375$$

$$\text{Gini}(\text{Sports}) = 1 - (8/8)^2 - (0/8)^2 = 0$$

$$\text{Gini}(\text{Luxury}) = 1 - (1/8)^2 - (7/8)^2 = 0,219$$

$$\text{Gini}(\text{Car Type}) = (4/20) * 0,375 + (8/20) * 0 + (8/20) * 0,219 = 0,163$$

$$\text{Entropi}(\text{Family}) = - (1/4) \log_2(1/4) - (3/4) \log_2(3/4) = 0,811$$

$$\text{Entropi}(\text{Sports}) = - (8/8) \log_2(8/8) - (0/8) \log_2(0/8) = 0$$

$$\text{Entropi}(\text{Luxury}) = - (1/8) \log_2(1/8) - (7/8) \log_2(7/8) = 0,544$$

$$\text{Entropi} = (4/20) * 0,811 + (8/20) * 0 + (8/20) * 0,544 = 0,379$$

E)

$$\text{Gini}(\text{Small}) = 1 - (3/5)^2 - (2/5)^2 = 0,48$$

$$\text{Gini}(\text{Medium}) = 1 - (3/7)^2 - (4/7)^2 = 0,4898$$

$$\text{Gini}(\text{Large}) = 1 - (2/4)^2 - (2/4)^2 = 0,5$$

$$\text{Gini}(\text{Extra Large}) = 1 - (2/4)^2 - (2/4)^2 = 0,5$$

$$\text{Gini} = (5/20)*0,48 + (7/20)*0,4898 + (4/20)*0,5 + (4/20)*0,5 = 0,491$$

$$\text{Entropi}(\text{Small}) = - (3/5)\log_2(3/5) - (2/5)\log_2(2/5) = 0,971$$

$$\text{Entropi}(\text{Medium}) = - (3/7)\log_2(3/7) - (4/7)\log_2(4/7) = 0,985$$

$$\text{Entropi}(\text{Large}) = - (2/4)\log_2(2/4) - (2/4)\log_2(2/4) = 1$$

$$\text{Entropi}(\text{Extra Large}) = - (8/8)\log_2(8/8) - (0/8)\log_2(0/8) = 0$$

$$\text{Entropi} = (5/20)*0,971 + (7/20)*0,985 + (4/20)*1 + (4/20)*0 = 0,788$$

F)

Gender, Car Type veya Shirt Size hangi özellik daha iyidir?

Gini ve entropisi en düşük olan özellik tercih edilir.

Gender = Gini :0.48 Entropi :0.97

Car type = Gini :0.163 Entropi :0.379 Car Type özelliği en iyi sonucu verir.

Shirt Size = Gini : = 0,491 Entropi : 0,788

G)

Customer ID en düşük Gini'ye sahip olmasına rağmen öznitelik testi koşulu olarak neden

Kullanılmaması gerektiğini açıklayın.

Çünkü tüm ID'ler benzersizdir ve bu özniteliği kullanarak hiçbir bölme yapılamaz. Bu nedenle,

Öznitelik testi koşulu olarak kullanılmamalıdır. Kullanıldığında modelin performansı düşer çünkü genelleme yapılamaz her bir id primary özelliktir böylece modelden verim alınamaz.

İkili sınıflandırma problemi için aşağıdaki tabloda gösterilen eğitim örneklerini ele alalım. Örneğin,

(a) Eğitim örneklerinin genel koleksiyonu için (geni endeksi) ve

(b) Customer ID formatı için Giri endeksini ve entropiyi hesaplayın.

(c) Cinsler dağılışı için Gini endeksimi ve entropiyi hesaplayınız.

(ii) Çok yollu bölme (multiway split) kullanılarak Car Type değişkeni için Gini endeksinin ve entropiyi hesaplayın.

(e) Çok yollu bölünmeyi kullanarak Shirt Size dağılışı için Gini endeksinin ve entropiyi hesaplayın.

(f) Gender, Car Type veya Shun Siao hangı ile ilgili daha yıldı?

(g) Customer ID en büyük gün'ye sahip olmasına rağmen dantelik testi kapulu olarak neden kullanılmaması gerektiğini açıklayın.

$$d) G_{\text{min}} = 1 - \frac{P(0/\text{OK})}{P(0/\text{OK})} = \frac{P(1/\text{OK})}{P(1/\text{OK})} = 1 - \frac{1/20}{1/20} = 1 - \frac{1}{20} = \frac{19}{20} = 0,95$$

$$Entrop = -10/20 \log_2(10/20) - 10/20 \log_2(10/20) = 1$$

$$b) G_{ii} = 1 - (1/1)^2 - (0/1)^2 = 0$$

$$\text{Entropi} = -1/1 \log_2(1/1) - (0/1) \log_2(0/1) = 0$$

c) $M \cdot G_{MI} = 1 - P(G_0/M)^2 - P(G_1/M)^2 =$
 $= 1 - \left(\frac{6}{10}\right)^2 - \left(\frac{4}{10}\right)^2 = 1 - \frac{36+16}{100} = \frac{48}{100} = 0,48$

$$F_{\text{crit}} = 1 - P(0|F) - P(1|F) = 1 - (4/10) - (6/10) = 0,48$$

$$r = \frac{10}{20} * 0,48 + \frac{10}{20} * 0,48 = 0,48$$

$$\text{Entropi } M = -\frac{6}{10} \log_2(6/10) - \frac{4}{10} \log_2(4/10) = 0,917$$

$$\text{Entropi } F = -\frac{4}{10} \log_2 \left(\frac{4}{10} \right) - \frac{6}{10} \log_2 \left(\frac{6}{10} \right) = 0,97$$