

Veri Modeliçilięi Ödev Seti 3

Musa Akçü2
181900060

Soru 1-9) OneR algoritmasında Veri setindeki her bir sınıf için kural yazılır. Frekans tablosuna göre en iyi değeriye kim veriyorsa o kural baz alınarak devam edilir.

Verilen table için frekans tablolarını oluşturalım.

Table 1		Kredi verme	
		Evet	Hayır
Yaş	30'un altında	2	3
	30'larında	4	0
	Yaşlı	3	2

IF yaş < 30 then Hayır
IF 30 ≤ yaş < 40 then Evet
IF yaş > 40 then Evet

Table 2		Kredi verme	
		Evet	Hayır
Gelir	Yüksek	3	1
	Orta	4	2
	Düşük	2	2

IF gelir == Yüksek then Evet
IF gelir == Orta then Evet
IF gelir == düşük then Hayır

Table 3		Kredi verme	
		Evet	Hayır
Öğrenci	Evet	3	3
	Hayır	6	2

IF öğrenci == evet then Hayır
IF öğrenci == hayır then Evet

Table 4		Kredi verme	
		Evet	Hayır
Kredi Durumu	Uygun	3	3
	Mükemmel	6	2

IF krediD == uygun then Hayır
IF krediD == mükemmel then Evet

Table 1	Confusion Matrix		Kredi verme	
	Koristelik matrisi		Evet	Hayır
			Evet	Hayır
OneR		Evet	7	2
		Hayır	2	3

	E	H
E		
H		

Positive Predictive Value (Pozitif Beklenen Değer) = $7/(7+2) = 0,78$
Negative Predictive Value (Negatif Beklenen Değer) = $3/(3+2) = 0,60$
Sensitivity = $7/(7+2) = 0,78$
Specificity = $3/(3+2) = 0,60$

$$Accuracy = (7+3)/(7+2+2+3) = \underline{\underline{0,71}}$$

Table 2	Confusion Matrix	Kredt värme	
One R	Erst	7	3
	Höger	2	2

	E	H
E		
H		

$$\text{Positive Predictive Value} = 7/(7+3) = 0,70$$

$$\text{Negative Predictive Value} = 2/(2+2) = 0,50$$

$$\text{Sensitivity} = 7/(7+2) = 0,78$$

$$\text{Specificity} = 2/(2+3) = 0,40$$

$$\text{Accuracy} = (7+2)/(7+3+2+2) = \underline{\underline{0,64}}$$

Table 3	Confusion Matrix	Kredt värme	
One R	Erst	3	3
	Höger	6	2

	E	H
E		
H		

$$\text{Positive Predictive Value} = 3/(3+3) = 0,50$$

$$\text{Negative Predictive Value} = 2/(2+6) = 0,25$$

$$\text{Sensitivity} = 3/(3+6) = 0,33$$

$$\text{Specificity} = 2/(2+3) = 0,40$$

$$\text{Accuracy} = (3+2)/14 = \underline{\underline{0,36}}$$

Table 4	Confusion Matrix	Kredt värme	
One R	Erst	6	0
	Höger	3	5

	E	H
E		
H		

$$\text{Positive Predictive Value} = 6/(6+0) = 1$$

$$\text{Negative Predictive Value} = 5/(5+3) = 0,63$$

$$\text{Sensitivity} = 6/(6+3) = 0,66$$

$$\text{Specificity} = 5/(5+0) = 1$$

$$\text{Accuracy} = 6+5/14 = \underline{\underline{0,78}}$$

Korsillik matrislerinden de gövüleceği üzere veri setindeki kredi derecesi sınıfının OneR'a göre tutarlılığı, tahmin edilebilirlik, diğer sınıflarından daha yüksektir. Bu nedenle karar tablosu olarak onunkini kullanacağız. %75'lik bir öğrenme gerektirecek. %12'lik hata payı olacak.

Karar Yapısı { IF krediDerecesi = uygun then HAYIR
IF krediDerecesi = mükemmel then EVET

b) Bu örnekte OneR algoritması hesaplamaları yaparak kredi derecesi ile ilgili karar için diğer eksik veri bilgilerini görmeyi gerektirir. Görmeyi gerektiren metodu kullanırız.

Yas Ortalama	Cinsiyet Ortalama	Öğrenim Ortalama	KD Ortalama	KD = uygun ise kredi vermek = HAYIR
?	Yüksek	Hayır	Mükemmel	KD = mükemmel ise kredi vermek = EVET
Yeni	?	?	Uygun	KD = uygun ise kredi vermek = HAYIR

Soru 2-a)

Verilen veri setindeki yaş değerlerine baktığımızda ve bunları sınıflandırdığımızda şöyle bir sonuç elde ediyoruz.

15 - 19 - 24 - 27 - 28 - 29 | 32 - 33 - 45 - 50 - 50 | 57 - 57 - 57 - 62 - 89
Yaşlar Yaşlar

< 30

30 ≥ x > 51

≥ 51

Bu dağılıma baktığımızda, çok net bir şekilde belli yaşlarda ve yaş gruplarında yığılma var. Bu yığılmanın analizini görmek için En büyük yaş ile en küçük yaş arasındaki fark ((89-15)/3) yapabiliriz. Grubu 3'e bölmek isterseniz 30'den aşağıda, 30 ile 51 arasında ve 51'den büyük şekilde yapabiliriz.

Yaş	Çeşit	Öğrenim	Kredi Notu	Kredi Verme
Az	Y	H	U	H
Orta	Y	E	M	E
Orta	Y	H	U	E
Yüksek	O	H	U	H
Yüksek	D	E	U	H
Yüksek	D	E	M	E
Az	D	E	M	E
Az	O	H	U	H
Az	D	E	U	H
Yüksek	O	H	M	H
Orta	O	E	U	E
Az	O	E	M	E
Orta	O	H	M	E
Az	Y	E	U	E
Orta	O	H	M	E
Yüksek	Y	H	U	E
Yüksek	Y	H	U	H

Her biri için frekans tablosu oluşturuz.

		Kredi Verme	
		E	H
Yaş	Az	3	3
	Orta	5	0
	Yüksek	2	4

		Kredi Verme	
		E	H
Çeşit	Y	4	2
	D	2	2
	O	4	3

		Kredi Verme	
		E	H
Öğrenim	E	6	2
	H	4	5

		Kredi Verme	
		E	H
Kredi Notu	U	4	6
	M	6	1

Yaş

IF yaş = az then HAYIR
IF yaş = orta then EVET
IF yaş = yüksek then HAYIR

Konsilik matrisi		k.v.	
		E	H
kval1	Evet	5	0
	Hayır	5	7

$$\text{Accuracy} = (5+7) / (5+0+5+7)$$

$$A = 0,71 \quad \text{Sensitivity} = 0,50$$

Coşkun

IF gelir = Yüksek then EVET
IF gelir = orta then EVET
IF gelir = düşük then HAYIR

Konsilik matrisi		k.v.	
		E	H
kval2	E	8	5
	H	2	2

$$\text{Accuracy} = (8+2) / (8+5+2+2)$$

$$A = 0,59$$

Öğrenci

IF öğrenci = evet then EVET
IF öğrenci = hayır then HAYIR

Konsilik matrisi		k.v.	
		E	H
kval3	E	6	2
	H	4	5

$$\text{Accuracy} = (6+5) / (6+2+4+5)$$

$$A = 0,65$$

Kredi notu

IF kredi notu = uygun then HAYIR
IF kredi notu = mük. then EVET

Konsilik matrisi		k.v.	
		E	H
kval4	E	6	1
	H	6	6

$$\text{Accuracy} = (6+6) / (6+1+4+6)$$

$$A = 0,71$$

$$\text{Sensitivity} = 0,60$$

Bu durumda yaş sınıfı ile kredi notu sınıfının doğruluk değerleri eşit çıktı. Hassasiyet değerlerini karşılaştırarak kredi notu sınıfı diğerine daha sağlıklı. Bu nedenle yine 1. Sınıftaki gibi onun karar yapısını kullanıyoruz.

b) Eksik değerler için bu seçer ağırlama yapmak kolay. Tüm örneklemlerin yaş değerleri sayısal değer olduğundan ortalamalarını alabiliriz. Hepsinin ortalamasını eksik değer kısmına atayabiliriz.

yaş	gelişim	öğrenme	kredi notu	
55 (orta)	orta	Hayır	Uygun	HAYIR
48 (ortalama) (orta)	Yüksek	Hayır	Uygun	HAYIR
79 (yaşlı) (yüksek)	?	?	Mükemmel	Evet