**FİNAL SORULARI**

1. Veri Madenciliği araçlarından olan WEKA, aşağıdaki programlama dillerinden hangisi ile geliştirilmiştir.?
   1. Matlab
   2. C#
   3. Python
   4. \* Java
   5. C++
2. Veri ambarı içeriğinin neler olduğunu belirten, kullanılan verinin yapısını ortaya koyan veri ambarının bileşeni aşağıdakilerden hangisidir.? (Quiz-1 8.soru)
   1. Düşük düzeyde özetlenmiş veri
   2. Ayrıntı veri
   3. Yüksek düzeyde özetlenmiş veri
   4. \* Metadata
   5. Eski ayrıntı veri

|  |  |  |  |
| --- | --- | --- | --- |
|  | GERÇEK | | |
| TAHMİN |  | **POZİTİF** | **NEGATİF** |
| **POZİTİF** | 12  True pozitif | 3  False pozitif |
| **NEGATİF** | 5  false negatif | 9  true negatif |

1. Yukarıda verilen karışıklık matrisine göre, Özgüllük(Specificity) değeri aşağıdakilerden hangisidir? (özgünlük = tn/(tn+fp))
   1. 12/15
   2. 12/17
   3. 12/31
   4. 9/14
   5. \* 9/12
2. Bir veri setine GİNİ algoritması uygulanarak karar ağacı oluşturulmak istenmektedir. Gini(j) aday bölünmesinin Gini(sol) değeri 0.51 ve Gini(sağ) değeri 0.57 olarak elde edilmiştir. Buna göre Gini(j) değeri aşağıdakilerden hangisi olabilir?
   1. 0.71
   2. 0.48
   3. 0.66
   4. 0.56
   5. 0.61

|  |  |
| --- | --- |
| **İşlemler** | **Satın alınan ürün listesi** |
| T1 | I1,I2,I5 |
| T2 | I2, I4 |
| T3 | I2, I3,I5 |
| T4 | I1, I2, I4 |
| T5 | I1, I3 |
| T6 | I2, I3 |
| T7 | I1, I3 |
| T8 | I1, I2, I3, I5 |

1. Yukarıda verilen verilere göre I5->I2,I3 birliktelik kuralına ait güven değeri aşağıdakilerden hangisidir?
   1. 0,5
   2. \* 0,66
   3. 0,33
   4. 0,4
   5. 0,2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HAVA | ISI | NEM | RÜZGAR | OYUN |
| GÜNEŞLİ | SICAK | YÜKSEK | KUVVETLİ | HAYIR |
| GÜNEŞLİ | SICAK | YÜKSEK | HAFİF | EVET |
| BULUTLU | ILIK | YÜKSEK | KUVVETLİ | HAYIR |
| YAĞMURLU | ILIK | NORMAL | KUVVETLİ | HAYIR |
| YAĞMURLU | SOĞUK | NORMAL | HAFİF | EVET |
| YAĞMURLU | SOĞUK | NORMAL | KUVVETLİ | HAYIR |

1. Yukarıda görülen veri setine sınıflandırma işlemi uygulanmış ve elde edilen karar ağacında ISI niteliğinin hiç kullanılmadığı görülmüştür. Buna göre sınıflandırma işleminde aşağıdaki algoritmalardan hangisi kesinlikle *kullanılmamıştır*?
   1. ID3
   2. Twoing
   3. En uzak komşu
   4. Gini
   5. C4.5
2. Birliktelik kuralında, X= destek(A,B)-destek(A)\*destek(B) formülü ile hesaplanan değer aşağıdakilerden hangisidir? ( quiz-2 5.soru)
   1. Lift
   2. Conviction
   3. Confidence
   4. Leverage ?
   5. coverage

|  |  |  |  |
| --- | --- | --- | --- |
|  | GERÇEK | | |
| TAHMİN |  | **POZİTİF** | **NEGATİF** |
| **POZİTİF** | 12  TP | 3  FP |
| **NEGATİF** | 5  FN | 9  TN |

1. Yukarıda verilen karışıklık matrisine göre test verilerinin kaç tanesi Pozitif sınıfına aittir?
   1. 17
   2. 9
   3. 12
   4. 21
   5. \* 15 (TP+FP)

|  |  |
| --- | --- |
| **İşlemler** | **Satın alınan ürün listesi** |
| T1 | I1,I2,I5 |
| T2 | I2, I4 |
| T3 | I2, I3,I5 |
| T4 | I1, I2, I4 |
| T5 | I1, I3 |
| T6 | I2, I3 |
| T7 | I1, I3 |
| T8 | I1, I2, I3, I5 |

1. Yukarıda verilen tabloda, 8 müşterinin satın aldığı 5 farklı ürün listelenmektedir. Bu veri setine MinSup=0.2 değeri için Apriori algoritması uygulandığında, 3 elemanlı yoğun nesne kümesinin (L3) eleman sayısı kaç olur? (Quiz-3 8.soru ile benzer)
   1. 1
   2. 2
   3. 0
   4. 4
   5. 3
2. Bir veri setine GİNİ algoritması uygulanarak karar ağacı oluşturulmak istenmektedir. Gini(j) aday bölünmesinin Gini(sol) değeri 0.51 ve Gini(sağ) değeri 0.57 olarak elde edilmiştir. Buna göre Gini(j) değeri aşağıdakilerden hangisi olabilir? ( final 4.soru)
   1. 0.71
   2. 0.48
   3. 0.66
   4. 0.56
   5. 0.61

|  |  |
| --- | --- |
| **İşlemler** | **Satın alınan ürün listesi** |
| T1 | I1,I2,I5 |
| T2 | I2, I4 |
| T3 | I2, I3,I5 |
| T4 | I1, I2, I4 |
| T5 | I1, I3 |
| T6 | I2, I3 |
| T7 | I1, I3 |
| T8 | I1, I2, I3, I5 |

1. Yukarıda verilen verilere göre I4-> I1 birliktelik kuralına ait LİFT değeri aşağıdakilerden hangisidir?
   1. 1.06
   2. 0.53
   3. 0.8
   4. 1.33
   5. 1.77
2. X=[1 2 4 7 12 20 35] verisine ait dendrogram, en yakın komşu algoritması ile oluşturulmuştur. Bu dendrogram kullanılarak veriler 2 kümeye ayrılmak istendiğinde, eleman sayısı en fazla olan kümenin eleman sayısı kaç olur? (Quiz-3 3.soru ve Quiz-2 8.soru ile benzer)
   1. 2
   2. 5
   3. 3
   4. 6
   5. 4

|  |  |
| --- | --- |
| **İşlemler** | **Satın alınan ürün listesi** |
| T1 | I1,I2,I5 |
| T2 | I2, I4 |
| T3 | I2, I3,I5 |
| T4 | I1, I2, I4 |
| T5 | I1, I3 |
| T6 | I2, I3 |
| T7 | I1, I3 |
| T8 | I1, I2, I3, I5 |

1. Yukarıda verilen verilere göre I2-> I5 birliktelik kuralına ait LİFT değeri aşağıdakilerden hangisidir?
   1. 1.33
   2. 1.06
   3. 0.53
   4. 1.77
   5. 2
2. Bir veri setinde verileri normalize etmek aşağıdakilerden hangisi olarak adlandırılır?

(quiz-2 2.soru)

* 1. Sıkıştırma
  2. Temizleme
  3. Bütünleştirme
  4. İndirgeme
  5. \* Dönüştürme

1. X=[3 7 13 4 5 10} verilerinin standart sapma değeri aşağıdakilerden hangisidir?
   1. -2.14
   2. 12.41
   3. 0.00
   4. \* 3.84
   5. 7.00

<https://standart-sapma.hesaplama.net/hesaplama.do>

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BAŞVURU | EĞİTİM | YAŞ | CİNSİYET | KABUL |
| 1 | ORTA | YAŞLI | ERKEK | EVET |
| 2 | İLK | GENÇ | ERKEK | HAYIR |
| 3 | YÜKSEK | ORTA | KADIN | HAYIR |
| 4 | ORTA | ORTA | ERKEK | EVET |
| 5 | İLK | ORTA | ERKEK | EVET |
| 6 | YÜKSEK | YAŞLI | KADIN | EVET |
| 7 | İLK | GENÇ | KADIN | HAYIR |

1. Şekilde verilen veri seti için Twoing algoritması ile karar ağacı bulunmak isteniyor. “orta, ilk🡨> yüksek” aday bölünmesinin P(evet|tsol) değeri aşağıdakilerden hangisidir?
   1. 0.75
   2. 0.50
   3. 0.60
   4. 0.33
   5. 0.25

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***A*** | ***B*** | ***C*** | ***D*** | ***E*** | ***F*** | ***G*** | ***H*** | ***İ*** | ***j*** |
| **X1** | 3 | 4 | 3 | -2 | -5 | -7 | 8 | 13 | 11 | 9 |
| **X2** | 5 | 8 | -6 | 4 | 5 | 1 | 2 | 7 | -4 | 2 |

1. Yukarıdaki tabloda 10 adet veriye ait x1 ve x2 değerleri verilmektedir. Bu veriler, k-ortalamalar kümeleme algoritması ile 5 kümeye ayrılmak istenmektedir. Başlangıç kümeleri, c1={a,b}, c2={c,d} c3={e,f} c4={g,h} c5={i,j} olarak belirlenmiştir. Algoritmanın birinci tekrarı (iterasyon) sonunda İ verisinin yeni kümesi aşağıdakilerden hangisi olur?
   1. C1
   2. C3
   3. C4
   4. C5
   5. C2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HAVA | ISI | NEM | RÜZGAR | OYUN |
| GÜNEŞLİ | SICAK | YÜKSEK | KUVVETLİ | HAYIR |
| GÜNEŞLİ | SICAK | YÜKSEK | HAFİF | EVET |
| BULUTLU | ILIK | YÜKSEK | KUVVETLİ | HAYIR |
| YAĞMURLU | ILIK | NORMAL | KUVVETLİ | HAYIR |
| YAĞMURLU | SOĞUK | NORMAL | HAFİF | EVET |
| YAĞMURLU | SOĞUK | NORMAL | KUVVETLİ | HAYIR |

1. Yukarıda verilen veri setini dikkate alarak P(ISIsıcak\OYUNhayır) olasılık değerini hesaplayınız?
   1. 1.00
   2. 0.25
   3. 0.33
   4. 0.50
   5. 0.75

|  |  |  |  |
| --- | --- | --- | --- |
|  | GERÇEK | | |
| TAHMİN |  | **POZİTİF** | **NEGATİF** |
| **POZİTİF** | 12  TP | 3  FP |
| **NEGATİF** | 5  FN | 9  TN |

1. Yukarıda verilen karışıklık matrisine göre, sınıflandırıcı kaç tane kaydı Negatif olarak tahmin etmiştir?
   1. 9
   2. \* 14 ( FN + TN )
   3. 12
   4. 17
   5. 15

|  |  |  |  |
| --- | --- | --- | --- |
|  | GERÇEK | | |
| TAHMİN |  | **POZİTİF** | **NEGATİF** |
| **POZİTİF** | 12  TP | 3  FP |
| **NEGATİF** | 5  FN | 9  TN |

1. Yukarıda verilen Karışıklık matrisine göre, Kesinlik(precision) değeri aşağıdakilerden hangisidir? Kesinlik=(tp/(tp+fp))
   1. 9/14
   2. 12/17
   3. 12/31
   4. 9/12
   5. \* 12/15

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BAŞVURU | EĞİTİM | YAŞ | CİNSİYET | KABUL |
| 1 | ORTA | YAŞLI | ERKEK | EVET |
| 2 | İLK | GENÇ | ERKEK | HAYIR |
| 3 | YÜKSEK | ORTA | KADIN | HAYIR |
| 4 | ORTA | ORTA | ERKEK | EVET |
| 5 | İLK | ORTA | ERKEK | EVET |
| 6 | YÜKSEK | YAŞLI | KADIN | EVET |
| 7 | İLK | GENÇ | KADIN | HAYIR |

1. Şekilde verilen veri seti için Twoing algoritması ile karar ağacı bulunmak isteiyor. “orta, yüksek🡨-> ilk” aday bölünmesinin P(Hayır|tsol) değeri aşağıdakilerden hangisidir?
   1. 0.25
   2. 0.33
   3. 0.50
   4. 0.75
   5. 0.60

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BAŞVURU | EĞİTİM | YAŞ | CİNSİYET | KABUL |
| 1 | ORTA | YAŞLI | ERKEK | EVET |
| 2 | İLK | GENÇ | ERKEK | HAYIR |
| 3 | YÜKSEK | ORTA | KADIN | HAYIR |
| 4 | ORTA | ORTA | ERKEK | EVET |
| 5 | İLK | ORTA | ERKEK | EVET |
| 6 | YÜKSEK | YAŞLI | KADIN | EVET |
| 7 | İLK | GENÇ | KADIN | HAYIR |

1. Şekilde verilen veri seti için Twoing algoritması ile karar ağacı bulunmak isteniyor. “orta, yüksek🡨-> ilk” aday bölünmesinin P(Hayır|tsağ) değeri aşağıdakilerden hangisidir?
   1. 0.25
   2. 0.50
   3. 0.33
   4. 0.60
   5. 0.75
2. K-En Yakın Komşu algoritmasında ağırlıklı uzaklık kullanmak için uzaklığa (d) aşağıdaki işlemlerden hangisi uygulanır? ( quiz-2 14.soru)
   1. D^2/2
   2. D^2
   3. 1-d
   4. \* 1/d^2
   5. (Kök içinde d)
3. X=[1 2 4 7 12 20 35] verisine ait dendrogram, en yakın komşu algoritması ile oluşturulmuştur. Bu dendrogram kullanılarak veriler 3 kümeye ayrılmak istendiğinde, eleman sayısı en fazla olan kümenin eleman sayısı kaç olur?
   1. 2
   2. 5
   3. 3
   4. 6
   5. 4
4. Bir veri setine GİNİ algoritması uygulanarak karar ağacı oluşturulmak istenmektedir. Gini(j) aday bölünmesinin Gini(sol) değeri 0.(resimde gözükmüyor) ve Gini(sağ) değeri 0.70 olarak elde edilmiştir. Buna göre Gini(j) değeri aşağıdakilerden hangisi olabilir?
   1. 0.56
   2. 0.66
   3. 0.48
   4. 0.61
   5. 0.71

|  |  |
| --- | --- |
| **İşlemler** | **Satın alınan ürün listesi** |
| T1 | I1,I2,I5 |
| T2 | I2, I4 |
| T3 | I2, I3,I5 |
| T4 | I1, I2, I4 |
| T5 | I1, I3 |
| T6 | I2, I3 |
| T7 | I1, I3 |
| T8 | I1, I2, I3, I5 |

1. Yukarıda verilen tabloda, 8 müşterinin satın aldığı 5 farklı ürün listelenmektedir. Bu veri setine MinSup=0.1 değeri için Apriori algoritması uygulandığında, 3 elemanlı yoğun nesne kümesinin (C3) eleman sayısı kaç olur?
   1. 1
   2. 2
   3. 0
   4. 4
   5. 3
2. Aşağıdakilerden hangisi, Birliktelik Kuralına ait bir metriktir?
   1. Kazanç
   2. Uygunluk değeri
   3. Öklid Uzaklığı
   4. Gini
   5. \* Destek
3. Zaman içerisinde olabildiğince birikmiş verilerin oluşturduğu bir veri yığınlarına ne denir?
   1. \* Veri ambarı
   2. OLTP
   3. OLAP
   4. Veri Madenciliği
   5. Veri Küpü
4. Aşağıdakilerden hangisi, Birliktelik Kuralında A🡪B kuralının güven değeri olabilir?

(not güven değeri = 0 ile 1 arasındadır)

* 1. \* 0.333
  2. -0.556
  3. 2.55
  4. 1.53
  5. -0.71

1. Aşağıdaki algoritmalardan hangisi entropiye dayalı bölümlemeyi kullanan bir sınıflandırma algoritmadır? ( quiz-2 7.soru)

(not = ID3 ve C4.5 algoritmaları entropi tabanlı algoritmalardır.)

* 1. Bayes
  2. Apriori
  3. \* C4.5
  4. SVM
  5. YSA

1. Meyve alan müşterilerin Meyse Suyu da aldığını gösteren kural, aşağıdaki birliktelik kuralı türlerinden hangisine aittir? (Quiz-2 9.soru ve Quiz-3 4.soru )

(Cevap = sıralı örüntü)

* 1. Boolean Birliktelik Kuralı
  2. Hiyerarşik Birliktelik Kuralı
  3. Ağırlıklandırılmış Birliktelik Kuralı
  4. \* Hiçbiri
  5. Periyodik Birliktelik Kuralı

1. Aşağıdakilerden hangisi Veri indirgeme yöntemlerinden biri değildir? ( quiz-2 13.soru)
   1. Veri küpü
   2. Sıkıştırma
   3. Örnekleme
   4. Genelleme
   5. \* Bütünleştirme

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **X1** | **X2** | **X3** | **SINIF** |
| **1** | C | 7 | 6 | C |
| **2** | 2 | 2 | 2 | B |
| **3** | 5 | 6 | 7 | C |
| **4** | 3 | 5 | 8 | C |
| **5** | 1 | 9 | 2 | D |
| **6** | 4 | 3 | 2 | A |
| **7** | 5 | 2 | 6 | B |
| **8** | 6 | 1 | 8 | E |

1. X={8,1,6} Değerlerine sahip bir veri K-En yakın komşu (KNN) algoritması ile sınıflandırıldığında k=1 değeri için X verisinin sınıfı aşağıdakilerden hangisi olur?
   1. B
   2. C
   3. D
   4. E
   5. A

|  |  |  |  |
| --- | --- | --- | --- |
|  | GERÇEK | | |
| TAHMİN |  | **POZİTİF** | **NEGATİF** |
| **POZİTİF** | 12  TP | 3  FP |
| **NEGATİF** | 5  FN | 9  TN |

1. Yukarıda verilen Karışıklık matrisine göre, sınıflandırıcının hatalı tahmin ettiği kayıt sayısı aşağıdakilerden hangisidir?
   1. 12
   2. 15
   3. \* 8 (FP + FN)
   4. 9
   5. 17
2. Bir veri setine GİNİ algoritması uygulanarak karar ağacı oluşturulmak istenmektedir. Gini(j) aday bölünmesinin Gini(sol) değeri 0.75 ve Gini(sağ) değeri 0.96 olarak elde edilmiştir. Buna göre Gini(j) değeri aşağıdakilerden hangisi olabilir?
   1. 0.71
   2. 0.61
   3. 0.81
   4. 0.56
   5. 0.66

|  |  |  |  |
| --- | --- | --- | --- |
|  | GERÇEK | | |
| TAHMİN |  | **POZİTİF** | **NEGATİF** |
| **POZİTİF** | 12  TP | 3  FP |
| **NEGATİF** | 5  FN | 9  TN |

1. Yukarıda verilen karışıklık matrisine göre, Duyarlık( sensitivity) değeri aşağıdakilerden hangisidir? (DUYARLILIK= ( TP/ (TP+FN))
   1. 12/31
   2. 12/15
   3. 9/14
   4. \* 12/17
   5. 9/12

**QUİZ-2 SORULARI**

|  |  |
| --- | --- |
| **İşlemler** | **Satın alınan ürün listesi** |
| T1 | I1,I2,I5 |
| T2 | I2, I4 |
| T3 | I2, I3,I5 |
| T4 | I1, I2, I4 |
| T5 | I1, I3 |
| T6 | I2, I3 |
| T7 | I1, I3 |
| T8 | I1, I2, I3, I5 |

1. Yukarıda verilen verilere göre I1🡪I3 birliktelik kuralına ait güven değeri aşağıdakilerden hangisidir?
   1. \* 0.6
   2. 0.5
   3. 0.66
   4. 0.4
   5. 0.75
2. Bir veri setinde verileri normalize etmek aşağıdakilerden hangisi olarak adlandırılır?

( final 14.soru)

* 1. Sıkıştırma
  2. Temizleme
  3. Bütünleştirme
  4. İndirgeme
  5. \* Dönüştürme

|  |  |  |  |
| --- | --- | --- | --- |
|  | Öngörülen sınıf | | |
| Doğru sınıf |  | **Sınıf=1** | **Sınıf=0** |
| **Sınıf=1** | A  TP | B  FN |
| **Sınıf=0** | C  FP | D  TN |

A:TP B:FN C:FP D:TN

1. Verilen karışıklık matrisine göre doğruluğun değeri aşağıdakilerden hangisidir?

Doğruluk ( accuracy) = (tp+tn) / (tp+fn+tn+fp)

* 1. D/(D+C)
  2. \* (A+D)/(A+B+C+D)
  3. A/(A+C)
  4. D/(B+D)
  5. A/(A+B)

|  |  |
| --- | --- |
| **İşlemler** | **Satın alınan ürün listesi** |
| T1 | I1,I2,I5 |
| T2 | I2, I4 |
| T3 | I2, I3,I5 |
| T4 | I1, I2, I4 |
| T5 | I1, I3 |
| T6 | I2, I3 |
| T7 | I1, I3 |
| T8 | I1, I2, I3, I5 |

1. Yukarıda verilen verilere göre I1,I2 🡪I3 birliktelik kuralına ait lift değeri aşağıdakilerden hangisidir?
   1. 1.16
   2. 0.88
   3. 0.94
   4. 1.07
   5. 1.23
2. Birliktelik kuralında, X= destek(A,B)-destek(A)\*destek(B) formülü ile hesaplanan değer aşağıdakilerden hangisidir? ( final 7.soru)
   1. Lift
   2. Conviction
   3. Confidence
   4. \* Leverage
   5. Coverage

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HAVA | ISI | NEM | RÜZGAR | OYUN |
| GÜNEŞLİ | SICAK | YÜKSEK | KUVVETLİ | HAYIR |
| GÜNEŞLİ | SICAK | YÜKSEK | HAFİF | EVET |
| BULUTLU | ILIK | YÜKSEK | KUVVETLİ | HAYIR |
| YAĞMURLU | ILIK | NORMAL | KUVVETLİ | HAYIR |
| YAĞMURLU | SOĞUK | NORMAL | HAFİF | EVET |
| YAĞMURLU | SOĞUK | NORMAL | KUVVETLİ | HAYIR |

1. Yukarıda verilen veri setine corona virüsü bulaştığı için bazı kayıtlar görülememektedir. Ancak Gini(J) aday bölünmesinin Gini(sol) değeri 0.48 ve Gini(sağ) değerinin 0.55 olduğu bilinmektedir. Buna göre Gini(j) değeri aşağıdakilerden hangisi olabilir?
   1. 0.57
   2. 0.50
   3. 0.60
   4. 0.41
   5. 0.36
2. Aşağıdaki algoritmalardan hangisi entropiye dayalı bölümlemeyi kullanan bir sınıflandırma algoritmadır? ( final 30.soru)
   1. Bayes
   2. Apriori
   3. C4.5
   4. SVM
   5. YSA
3. X=[1 2 4 6 12 20] verisine ait dendrogram, en yakın komşu algoritması ile oluşturulmuştur. Bu dendrogram kullanılarak veriler 2 kümeye ayrılmak istendiğinde, eleman sayısı en fazla olan kümenin eleman sayısı kaç olur? (Quiz-3 3.soru ve final 12. Soru ile benzer)
   1. 3
   2. 0
   3. 4
   4. 1
   5. 2
4. Kola alan müşterilerin Cips de aldığını gösteren kural, aşağıdaki birliktelik kuralı türlerinden hangisine aittir? (Quiz-3 4.soru ve final 31.soru)
   1. Boolean Birliktelik Kuralı
   2. Hiyerarşik Birliktelik Kuralı
   3. Ağırlıklandırılmış Birliktelik Kuralı
   4. Periyodik Birliktelik Kuralı
   5. \* Hiçbiri

|  |  |
| --- | --- |
| **İşlemler** | **Satın alınan ürün listesi** |
| T1 | I1,I2,I5 |
| T2 | I2, I4 |
| T3 | I2, I3,I5 |
| T4 | I1, I2, I4 |
| T5 | I1, I3 |
| T6 | I2, I3 |
| T7 | I1, I3 |
| T8 | I1, I2, I3, I5 |

1. Yukarıda verilen veriler üzerinde Apriori algoritması çalıştırıldığında 2 elemanlı yoğun nesne kümesinin eleman sayısı 3 olarak bulunmuştur. Buna göre Algoritmada kullanılan Minimum Destek değeri aşağıdakilerden hangisidir?
   1. \* 0.25
   2. 0.2
   3. 0.40
   4. 0.35
   5. 0.45

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HAVA | ISI | NEM | RÜZGAR | OYUN |
| GÜNEŞLİ | SICAK | YÜKSEK | KUVVETLİ | HAYIR |
| GÜNEŞLİ | SICAK | YÜKSEK | HAFİF | EVET |
| BULUTLU | ILIK | YÜKSEK | KUVVETLİ | HAYIR |
| YAĞMURLU | ILIK | NORMAL | KUVVETLİ | HAYIR |
| YAĞMURLU | SOĞUK | NORMAL | HAFİF | EVET |
| YAĞMURLU | SOĞUK | NORMAL | KUVVETLİ | HAYIR |

1. Yukarıda verilen veri setine, ID3 algoritması uygulanarak Oyun sınıfını tahmin eden bir karar ağacı oluşturulmak istenmektedir. Oluşacak Karar ağacının kökönü, aşağıdaki niteliklerden hangisinin oluşturduğunu bulunuz?
   1. ISI
   2. OYUN
   3. HAVA
   4. RÜZGAR
   5. NEM

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HAVA | ISI | NEM | RÜZGAR | OYUN |
| GÜNEŞLİ | SICAK | YÜKSEK | KUVVETLİ | HAYIR |
| GÜNEŞLİ | SICAK | YÜKSEK | HAFİF | EVET |
| BULUTLU | ILIK | YÜKSEK | KUVVETLİ | HAYIR |
| YAĞMURLU | ILIK | NORMAL | KUVVETLİ | HAYIR |
| YAĞMURLU | SOĞUK | NORMAL | HAFİF | EVET |
| YAĞMURLU | SOĞUK | NORMAL | KUVVETLİ | HAYIR |

1. 6 kayıt bulunan yukarıda veri setine corona virüsü bulaştığı için, 2 niteliğine ait 2 şer tane veri görülememektedir. Ancak [ Yüksek,Düşük 🡨🡪 Normal] aday bölünmesinin Gini değeri ile Gİni(sol) değerinin eşit olduğu biinmektedir. Buna göre Gini(sağ) değeri aşağıdakilerden hangisidir?
   1. 0.37
   2. 0.48
   3. 0.44
   4. 0.51
   5. 0.32
2. Aşağıdakilerden hangisi Veri indirgeme yöntemlerinden biri değildir? ( final 32.soru)
   1. Veri küpü
   2. Sıkıştırma
   3. Örnekleme
   4. Genelleme
   5. Bütünleştirme
3. K-En Yakın Komşu algoritmasında ağırlıklı uzaklık kullanmak için uzaklığa (d) aşağıdaki işlemlerden hangisi uygulanır? ( final 23.soru)
   1. D^2/2
   2. D^2
   3. 1-d
   4. 1/d^2
   5. (Kök içinde d)

**QUİZ-3 SORULARI**

1. X1=[3 7 1 4 5 2] verisi ile X2=[2 6 11 4 8 1] verisi arasındaki Manhattan uzaklığını hesaplayınız?
2. Quiz1 veya Quiz2 nin herhangi birinde, sorular veya cevaplar elime ulaştı ve bunlardan faydalandım.
   1. Doğru
   2. Yanlış
3. X=[1 2 4 6 12 20] verisine ait dendrogram, en yakın komşu algoritması ile oluşturulmuştur. Bu dendrogram kullanılarak veriler 2 kümeye ayrılmak istendiğinde, eleman sayısı en fazla olan kümenin eleman sayısı kaç olur? (Quiz-2 8.soru ve final 12. Soru ile benzer)
   1. 3
   2. 0
   3. 4
   4. 1
   5. 2
4. “Bebek bezi” alan müşterilerin “Mama” da aldığını gösteren kural, aşağıdaki birliktelik kuralı türlerinden hangisine aittir? (Quiz-2 9.soru ve final 31.soru)
   1. Boolean Birliktelik Kuralı
   2. Hiyerarşik Birliktelik Kuralı
   3. Ağırlıklandırılmış Birliktelik Kuralı
   4. Periyodik Birliktelik Kuralı
   5. Sıralı Örüntüler
5. Bu ders için hazırlanan ders videolarını yeterince izlediğinize, ödevleri yapmaya çalıştığınıza ve yeterli emeği verdiğinize inanıyor musunuz? 1 ile 5 puan arasında bir puan veriniz.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***A*** | ***B*** | ***C*** | ***D*** | ***E*** | ***F*** | ***G*** | ***H*** | ***İ*** | ***j*** |
| **X1** | 3 | 4 | 3 | -2 | -5 | -7 | 8 | 13 | 11 | 9 |
| **X2** | 5 | 8 | -6 | 4 | 5 | 1 | 2 | 7 | -4 | 2 |

1. Yukarıda verilen veri setine ait dendrogram, minskowski uzaklık ölçütü ve en yakın komşu algoritması kullanılarak oluşturulmak istenmektedir. Uzaklık matrisleri hesaplanırken, aşağıdaki değerlerden hangisi elde edilir. ( Not: normal yollarla çözmeye çalışırsanız soru çok zaman alacaktır. Mantık yürüterek çözmeye çalışınız)
   1. 4.44
   2. 34.21
   3. 17.83
   4. 12.47
   5. 54.82
2. Elimizde 2 adet zar bulunmaktadır. Bu zarkardan birinde TEK sayılar( 1,3,5,7,9,11) sayılar diğerinde ise ÇİFT sayılar(2,4,6,8,10,12) bulunmaktadır. Rastgele seçilen bir zar 2 defa atıldığında gelen sayıların toplamın 10’dan küçük olduğunu bilindiğine göre bu rastgele seçilen zarın ÇİFT sayıların olduğu zar olma ihtimali nedir?
   1. 0.50
   2. 0.66
   3. 0.75
   4. 0.25
   5. 0.33

|  |  |
| --- | --- |
| **İşlemler** | **Satın alınan ürün listesi** |
| T1 | I1,I2,I5 |
| T2 | I2, I4 |
| T3 | I2, I3,I5 |
| T4 | I1, I2, I4 |
| T5 | I1, I3 |
| T6 | I2, I3 |
| T7 | I1, I3 |
| T8 | I1, I2, I3, I5 |

1. Yukarıda verilen tabloda, 8 müşterinin satın aldığı 5 farklı ürün listelenmektedir. Bu veri setine MinSup=0.2 değeri için Apriori algoritması çalıştırıldığında, 3 elemanlı yoğun nesne kümesinin (L3) eleman sayısının 4 olması için veri setine en az kaç tane daha ürün eklenmesi gerekir. (müşteriler en az kaç tane daha ürün alsa idi L3’te 4 kayıt bulunurdur?)

(final 9.soru ile benzer)

1. A 🡪 B birliktelik kuralı için, aşağıdakilerden hangisi diğerlerine göre A ile B arasındaki ilişkinin en güçlü olduğu lift değeridir?
   1. 2.5
   2. 0.87
   3. -3.33
   4. 0
   5. 0.33

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HAVA | ISI | NEM | RÜZGAR | OYUN |
| GÜNEŞLİ | SICAK | YÜKSEK | KUVVETLİ | HAYIR |
| GÜNEŞLİ | SICAK | YÜKSEK | HAFİF | EVET |
| BULUTLU | ILIK | YÜKSEK | KUVVETLİ | HAYIR |
| YAĞMURLU | ILIK | NORMAL | KUVVETLİ | HAYIR |
| YAĞMURLU | SOĞUK | NORMAL | HAFİF | EVET |
| YAĞMURLU | SOĞUK | NORMAL | KUVVETLİ | HAYIR |

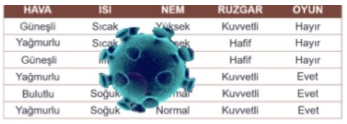
1. Yukarıda görülen veri seti üzerinde twoing algoritması kullanılarak ISI niteliğini tahmin edecek bir karar ağacı oluşturulmak istenmektedir. ISI niteliğinin tahmin edilebilmesi için 1. İterasyonda kaç farklı aday bölünme elde edilmektedir?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HAVA | ISI | NEM | RÜZGAR | OYUN |
| GÜNEŞLİ | SICAK | YÜKSEK | KUVVETLİ | HAYIR |
| GÜNEŞLİ | SICAK | YÜKSEK | HAFİF | EVET |
| BULUTLU | ILIK | YÜKSEK | KUVVETLİ | HAYIR |
| YAĞMURLU | ILIK | NORMAL | KUVVETLİ | HAYIR |
| YAĞMURLU | SOĞUK | NORMAL | HAFİF | EVET |
| YAĞMURLU | SOĞUK | NORMAL | KUVVETLİ | HAYIR |

1. Yukarıda verilen veri seti üzerinde birliktelik kuralı uygulandığında, elde edilen kurallar içerisinde lift değeri en yüksek olan kuralın lift değeri aşağıdakilerden hangisidir?
   1. 6
   2. 3
   3. 4
   4. 7
   5. 5
2. Aşağıdakilerden hangisi, Birliktelik Kuralına ait bir metriktir?
   1. Özgüllük
   2. F-ölçütü
   3. Minskowski
   4. Anma
   5. İlginçlik

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HAVA | ISI | NEM | RÜZGAR | OYUN |
| GÜNEŞLİ | SICAK | YÜKSEK | KUVVETLİ | HAYIR |
| GÜNEŞLİ | SICAK | YÜKSEK | HAFİF | EVET |
| BULUTLU | ILIK | YÜKSEK | KUVVETLİ | HAYIR |
| YAĞMURLU | ILIK | NORMAL | KUVVETLİ | HAYIR |
| YAĞMURLU | SOĞUK | NORMAL | HAFİF | EVET |
| YAĞMURLU | SOĞUK | NORMAL | KUVVETLİ | HAYIR |

1. Yukarıda görülen veri seti üzerinde Gini algoritması kullanılarak karar ağacı oluşturulmak istenmektedir. “Güneşli,Bulutlu 🡨🡪Yağmurlu” aday bölünmesine ait Gini değeri aşağıdakilerden hangisidir?
   1. 0.48
   2. 0.37
   3. 0.50
   4. 0.44
   5. 0.56
2. X=[2 3 4 6 11 12 15 16 17 24 25] verisi, k-ortalamalar algoritması ile k tane kümeye ayrılmak istenmektedir. Elaman sayısı en az olan küme ile en fazla olan kümenin eleman sayıları arasındaki farkın 3 olduğunu net olarak bilindiğine göre k değeri nedir?



1. Yukarıda görülen ve sadece 4 verinin etkilendiği bu virüslü veri seti, virüsten temizlemek istenmektedir. Virüsün etkilendiği bu 4 veri için virüs temizlendiğinde bu veriler aşağıdakilerden hangileri olur? ( not: Bu soruyu sınavdaki diğer sorulardan bağımsız olarak çözünüz.)
   1. ISI: soğuk, sıcak / NEM: Düşük,Yüksek
   2. ISI: sıcak, sıcak / NEM: normal, normal
   3. ISI: ılık, sıcak / NEM: Yüksek, normal
   4. ISI: sıcak, soğuk / NEM: Düşük,Yüksek
   5. \* ISI: giniılık, ılık / NEM: Yüksek, normal

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HAVA | ISI | NEM | RÜZGAR | OYUN |
| GÜNEŞLİ | SICAK | YÜKSEK | KUVVETLİ | HAYIR |
| GÜNEŞLİ | SICAK | YÜKSEK | HAFİF | EVET |
| BULUTLU | ILIK | YÜKSEK | KUVVETLİ | HAYIR |
| YAĞMURLU | ILIK | NORMAL | KUVVETLİ | HAYIR |
| YAĞMURLU | SOĞUK | NORMAL | HAFİF | EVET |
| YAĞMURLU | SOĞUK | NORMAL | KUVVETLİ | HAYIR |

1. Bayes sınıflandırıcı kullanarak yukarıda görülen veri seti ile sınıflandırma yapılacaktır. Birini dışarıda bırak çapraz doğrulama testi yöntemi kullanılarak elde edilecek başarım değeri aşağıdakilerden hangisi olabilir?
   1. 0
   2. 0.75

**QUİZ-1 SORULARI**

1. Aşağılardan hangisi veri ambarının özelliklerinden biri değildir?
   1. **Dinamik yapıya sahiptir +++**
   2. Bütünleşiktir
   3. Sadece okunabilir
   4. Zaman boyutu vardır
   5. Konuya yöneliktir
2. Aşağıdakilerden hangisi, Birliktelik Kuralında A🡪B kuralının güven değeri olabilir?
   1. **0.01 +++**
   2. -2.35
   3. 2
   4. 1.01
   5. -0,5

|  |  |  |  |
| --- | --- | --- | --- |
|  | Öngörülen sınıf | | |
| Doğru sınıf |  | **Sınıf=1** | **Sınıf=0** |
| **Sınıf=1** | A  TP | B  FN |
| **Sınıf=0** | C  FP | D  TN |

A:TP B:FN C:FP D:TN

1. Verilen karışıklık matrisine göre sınıflandırıcının hatalı tahmin ettiği kayıt sayısı kaçtır?
   1. C
   2. (C+D)
   3. (B+D)
   4. **(B+C) +++**
   5. D

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BAŞVURU | EĞİTİM | YAŞ | CİNSİYET | KABUL |
| 1 | ORTA | YAŞLI | ERKEK | EVET |
| 2 | İLK | GENÇ | ERKEK | HAYIR |
| 3 | YÜKSEK | ORTA | KADIN | HAYIR |
| 4 | ORTA | ORTA | ERKEK | EVET |
| 5 | İLK | ORTA | ERKEK | EVET |
| 6 | YÜKSEK | YAŞLI | KADIN | EVET |
| 7 | İLK | GENÇ | KADIN | HAYIR |

1. Şekilde verilen veri seti için Gini algoritması ile karar ağacı bulunmak isteniyor “ilk,Orta🡨🡪Yüksek” bölünmesinin Gini(sol) değeri aşağıdakilerden hangisidir?
   1. 0.44
   2. 0.38
   3. 0.50
   4. **0.48 +++**
   5. 0.32

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BAŞVURU | EĞİTİM | YAŞ | CİNSİYET | KABUL |
| 1 | ORTA | YAŞLI | ERKEK | EVET |
| 2 | İLK | GENÇ | ERKEK | HAYIR |
| 3 | YÜKSEK | ORTA | KADIN | HAYIR |
| 4 | ORTA | ORTA | ERKEK | EVET |
| 5 | İLK | ORTA | ERKEK | EVET |
| 6 | YÜKSEK | YAŞLI | KADIN | EVET |
| 7 | İLK | GENÇ | KADIN | HAYIR |

1. Şekilde verilen veri seti için Twoing algoritması ile karar ağacı bulunmak isteniyor. Psol=4/7 değeri aşağıdaki hangi aday bölünmeye aittir?
   1. **Erkek🡨🡪Kadın +++**
   2. Orta🡨🡪Genç, Yaşlı
   3. Yüksek, Genç 🡨🡪 İlk
   4. Orta 🡨🡪 İlk, Yüksek
   5. Orta, Yaşlı 🡨🡪 Genç

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BAŞVURU | EĞİTİM | YAŞ | CİNSİYET | KABUL |
| 1 | ORTA | YAŞLI | ERKEK | EVET |
| 2 | İLK | GENÇ | ERKEK | HAYIR |
| 3 | YÜKSEK | ORTA | KADIN | HAYIR |
| 4 | ORTA | ORTA | ERKEK | EVET |
| 5 | İLK | ORTA | ERKEK | EVET |
| 6 | YÜKSEK | YAŞLI | KADIN | EVET |
| 7 | İLK | GENÇ | KADIN | HAYIR |

1. Şekilde verilen veri seti için Twoing algoritması ile karar ağacı bulunmak isteniyor.

“Orta, Yaşlı🡨🡪Genç” bölünmesinin P(evet|tsağ) değeri aşağıdakilerden hangisidir?

* 1. **0 +++**
  2. 0.33
  3. 0.25
  4. 0.67
  5. 0.75

|  |  |  |  |
| --- | --- | --- | --- |
|  | Öngörülen sınıf | | |
| Doğru sınıf |  | **Sınıf=1** | **Sınıf=0** |
| **Sınıf=1** | A  True pozitif | B  False pozitif |
| **Sınıf=0** | C  False pozitif | D  True negatif |

A:TP B:FN C:FP D:TN

1. Test veri setinde 1 sınıfına ait toplam kaç kayıt bulunmaktadır?
   1. A+B+C
   2. A
   3. A+C
   4. A+B+C+D
   5. **A+B +++**
2. Veri ambarı içeriğinin neler olduğunu belirten, kullanılan verinin yapısını ortaya koyan veri ambarının bileşeni aşağıdakilerden hangisidir.? (final 2.soru)
   1. Düşük düzeyde özetlenmiş veri
   2. Ayrıntı veri
   3. Yüksek düzeyde özetlenmiş veri
   4. **Metadata +++**
   5. Eski ayrıntı veri
3. X=[1 2 4 7 12 20 33] verisine ait dendrogram, en yakın komşu algoritması ile oluşturulmak istenmektedir. Bu dendrogram kullanılarak veriler 2 kümeye ayrılmak istendiğinde, oluşan her iki kümenin eleman sayıları arasındaki fark kaç olur?
   1. 3
   2. 1
   3. 4
   4. 2
   5. **5 +++**
4. X=[3 7 13 4 5 10] verilerine max-min normalleştirme uygulanacaktır. Bu işlem yapılırken A şıkkında görülen verinin yeni değeri hesaplanırken, Hangi şıkta verilen değer ile farkının alınması gerekir?
   1. 13
   2. **3 +++**
   3. 7
   4. 4
   5. 10
5. X=[3 7 13 4 5 10] verilerinin standart sapma değeri aşağıdakilerden hangisidir?
   1. 7.00
   2. **3.84 +++**
   3. -2.14
   4. 12.41
   5. 0.00
6. Risk={kötü,kötü,kötü,iyi,iyi,iyi} veri grubunun H(T) Entropi değerini hesaplayınız?

**Cevap = 1**

1. Veri tabanında yer alan tutarsız ve hatalı verilere …… denir.

**Cevap = gürültü**

**Kaynaklar:**

<https://www.youtube.com/watch?v=Srsod3e_KF8&feature=youtu.be>

<https://www.youtube.com/watch?v=AkCwUiSlZHo&feature=emb_title>