## Apriori algoritması

Soru1.1: Aşağıdakilerden hangisi Apriori algoritmasının uygulama alanlarından değildir?

- a) Marketlerde ürünler arası ilişkileri tanımlamak
- b) Restoranlarda servis hızının artırılması için müşterilerin sipariş etme ihtimali olan ürünleri önceden tahmin ederek hazırlamak
- Depo içindeki hareketi ve taşıma miktarını azaltmak için birlikte sevk edilen ürünlerin yakın raflara yerleştirilmesi
- d) Belirtilere göre hastalık tahmini

Soru1.2: Literatürde birliktelik kuralı çıkaran değişik algoritmalar bulunmaktadır. Aşağıdakilerden hangisi birliktelik kuralı çıkarım algoritmaları içerisinde en fazla bilinen algoritmadır?

- a) Twiong
- b) Apriori
- c) Gini
- d) Prim

Soru1.3: Apriori algoritmasının her eleman için çalışmasını, bir arama algoritmasına benzetmek mümkündür. Algoritma, bu anlmada hangi arama yapısındadır?

- a) Sığ öncelikli arama
- b) Derin öncelikli arama
- c) Sınırlı derin öncelikli arama
- d) İkili arama

-----

VeriMadenciliğiSosyalEtkileri(Tehditler)

Soru2.1.Aşağıdakilerden hangisi veri madenciliğinde tehditleri önlemek için yapılan çalışmalardan değildir?

- A. Veri Koleksiyonu
- B. . Sonuç Kontrolü
- C. . Bilgi Paylaşımı
- D. .Giriş Kontrolü

.....

### rapidminer

Soru3.1Rapidminer'da Decision Tree algoritmasını sağlıklı olarak çalıştırmak için bir adet label değerine ihtiyaç vardır.Algoritmanın bu label üzerinden çalışması sonucunda ne olur?

- A) operatörün tanımını yazar
- B) Ekranı ikiye böler
- C) Split validation oluşturur
- D) Karar Ağacını döker
- E) Eğitilmiş veri setini görüntüler

Soru3.2 Aşağıdakilerden hangisi rapidminer(yale) programının özelliklerinden değildir?

A)3D ile verileri analiz etme			
B)yapay sinir ağları			
C)normalizasyon			
D)genetik			
E)C# diliyle oluşturulması			
Soru3.3 RapidMiner'da yeni bir proses başlatmak için doğru adımlar aşağıdakilerden hangisind doğru verilmiştir?			
A: File->New->			
B: File->Add->New			
C: File->Add New Item->New Process			
D: File->New->Add New Item			
E: File->New->Process			
Soru3.4 Aşağıdakilerden hangisi RapidMiner'da veri setini öğrenme ve test edilme kısımlarına ayıran			
operatörtür?			
A: Evalution			
B: Validation			
C: SplitValidation			
D: Training parce			
E: SplitTraining			
Soru3.5 Amerika'da bulunan YALE üniversitesi bilim adamları tarafından geliştirilen RapidMiner hangi programla dili ile geliştirilmiştir?			
A: C			
B: C#			
C: Java			
D: Cobol			
E: PL/I			
Zeki Veri Madenciliği			
Soru5.1-) Demir Emilimini Arttıran Durumlar hangisidir ?			
A) Midedeki asit salgısı B) C vitamini varlığı C) Proteinden zengin gıdalar			

- D) Vücuttaki demirin azalması
- E) Hepsi (cevap)

Soru5.2-) Anemi hastalığı teşhişi için hangi parametreler kullanılır?

```
I-) Hb
II-)Hct
III-)Mcv
IV-)Pro
A)I
B)I II
C)I III IV
D)I II III (cevap)
E)IV
```

Soru5.3-)K -means algoritması için hangisi yanlıştır?

- A) Eğer küme sayısı belirli değil ise deneme yoluyla en uygun sayı bulunur.
- B) K-means algoritması gürültü ve istisnaya karsı çok duyarlıdır
- C) Çakışan kümelerde iyi sonuç vermez.
- D) K-means algoritması sadece sayısal veriler ile kullanılabilmektedir.
- E) Hiçbiri (cevap)

Soru5.4 Aşağıdakilerden hangisi zeki veri madenciliği uygulama alanlarından değildir?

- A. Bilimsel ve Mühendislik Verileri
- B. Sağlık Verileri
- C. Doküman Verileri
- D. Kredi Verileri(cevap)

Soru5.5: Aşağıdakilerden hangisinde zeki veri madenciliği ve makine öğrenimi ilişkisinde kullanılan algoritmalar tam olarak verilmiştir?

- A. Yapay sinir ağları , karar ağacı tabanlı algoritmalar
- B. Karar ağacı tabanlı, doğrudan kural üreten algoritmalar(cevap)
- C. İstatistiksel algoritmalar, yapay sinir ağları
- D. Doğrudan kural üreten, yapay sinir ağları algoritmaları

-----

Veri Madenciliği Sosyal Etkileri (FIRSATLAR)

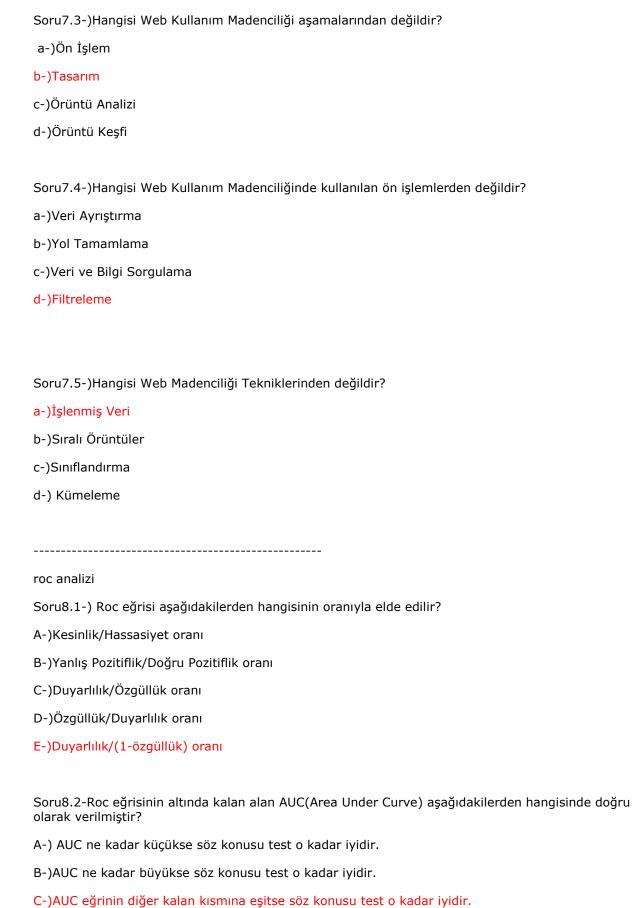
Soru6.1) Aşağıdakilerden hangisi Veri Madenciliğinin nedenlerinden değildir?

- A) Verilerdeki büyük artış
- B) Verilerle çok iç içe olmamız
- C) Büyük veri setlerinin otomatik analiz edilmesi ihtiyacı
- D) Veritabanı yönetim sistemlerinin yetersizliği

Soru6. 2) Aşağıdakilerden hangisi Veri Madenciliğinin faydalarından değildir?

- A) Geçmiş ve mevcut yapı analiz edilerek geleceğe yönelik tahminlerde bulunulabilir
- B) Müşterilerin kişisel bilgilerini başka kurum veya kuruluşlara pazarlamak
- C) Firmanın finansal yapısının, makro ekonomik değişmeler karşısındaki duyarlılığı ve oluşabilecek risklerin tespitinde kullanabilir

D)	Kurum teknik kaynaklarının en uygun şekilde kullanılmasını sağlamakta kullanılabilir			
Soru6.	3) Veri Madenciliği; aşağıdaki alanların hangisinde en az kullanım alanına sahiptir?			
A)	CRM (Müşteri ilişkileri yönetimi)			
B)	Bankacılık			
C)	Şans oyunları			
D)	Sağlık			
Soru6.	4) Aşağıdakilerden hangisi Veri Madenciliğinin faydalarından değildir?			
A)	A) Hile ve usulsüzlüklere sebebiyet vermesi			
B)	Maliyetlerin azaltılmasında yardımcı olması			
C)	Arge hizmetlerine yardımcı olması			
D)	Sahtekarlıkları tespit etme ve engellemede yardımcı olması			
	5) "Bankacılık", "Doğrudan pazarlama" ve "Eğlence" alanları kendi içerisinde sıralandığında tan en aza doğru Veri Madenciliği kullanım sırası aşağıdakilerden hangisi gibi olur?			
A)	Bankacılık > Eğlence > Doğrudan pazarlama			
B)	Doğrudan pazarlama > Eğlence > Bankacılık			
C)	Bankacılık > Doğrudan pazarlama > Eğlence			
D)	Eğlence > Bankacılık > Doğrudan pazarlama			
Web M	adenciliği			
Soru7.	1-)Hangisi Web Madenciliği tanımını karşılamaz?			
a-)Veri madenciliği tekniklerini kullanır				
b-)Kull	anıcıların web sitelerindeki davranışlarını inceler.			
c-) We	b belgelerinin ayıklanması ve analiz edilmesi			
d-) Erişim kayıtlarının içerdiği verilerin tamamı				
Soru7.	2-)Web Madenciliği kaç sınıfa ayrılır?			
a-)1				
b-)2				
c-)3				
d-)4				
•				



D-)AUC teste hiçbir zaman etki etmez.

E-)AUC için kesin bir şey ifade edilemez.		
birliktelik kuralları		
Soru9.1) Aşağıdakilerden hangisi birliktelik kurallarına ait değildir?		
a) Birliktelik kuralları analizi, yaygın olarak kullanılan veri madenciliği yöntemlerinden birisidir.		
b) Veriler arasındaki birlikteliklerin,ilişkilerin ve bağıntıların kurallar halinde bulunması işlemidir.		
c) Veri nesneleri arasındaki ilginç ilişkiler ve eş zamanlı gerçekleşen durumlar araştırılır.		
d) Birliktelik kuralı, eğer kuralın sol tarafı doğru ise sağ tarafının da doğru olmasını gerektirmez		
Soru9.2)Aşağıdakilerden hangisi apriori algoritması için doğrudur?		
a) Apriori Algoritması, bilgileri bir onceki adımdan aldığı icin "prior (onceki)" kelimesinden turetilmiştir.		
b) Apriori Algoritması geniş nesne kumelerinin ortaya cıkartılması işlemleri icin kullanılır.		
c) Bir nesne kumesinin geniş olarak adlandırılabilmesi icin ise o nesne kumesinin kullanıcı tarafından verilen minimum destek seviyesinin uzerinde bir destek seviyesine sahip olması gereki		
d)hepsi		
Soru9.3)Aşağıdakilerden hangisi FP-Growth algoritması için doğrudur?		
a) Diğer algoritmalara göre daha yüksek performans gosterir.		
b) Sıralama işlemi sayesinde destek değeri daha buyuk olan nesneler koke daha yakın olur		
c) Ağac oluşturulduktan sonra uzerinde Growth Algoritması calıştırılır.		
d)hepsi		
Soru9.4)Hangisi birliktelik algoritmalarından değildir?		
a)AIS		
b)SETC		
c)OCD		
d)FP-Growth		
Soru9.5) Birden fazla karşılaştırma elemanı yada özellik içeren birliktelik kurallarıdır bu tanıma en uvgun karşılık gelecek yapı hangisidir?		

- a) Tek boyutlu birliktelik kuralları
- b) boyutlar arası birliktelik kuralları
- Çok boyutlu birliktelik kuralları c)

d) hibrit birliktelik kuralları		
KARAR AĞAÇLARI		
Soru10.1 Aşağıdakilerden hangisi karar ağacı oluşturmanın temel adımlarındandır?		
A)Verilerin hazırlanması		
B)Verilerin teker teker ağaca uygulanarak sınıflandırmanın gerçekleştirilmesi		
C)Karar ağacı oluşturacak algoritmanın belirlenmesi		
D)Veri ambarının oluşturması		
E)Roc analizinin yapılması		
Soru10.2 Aşağıdakilerden hangisi karar ağacı oluşturma algoritmalarından değildir?		
A)ID3		
B)CART		
C)HUNT		
D)SLINQ		
E)SPRİNT		
Soru10.3 Aşağıdakilerden hangisi karar ağaçlarının avantajlarından değildir?		
A)Zaman ve yer karmaşıklığı öğrenme süresi örnekleri sayısına,nitelik sayısına ve oluşan ağacır yapısına bağlıdır		
B)Sürekli ve ayrık nitelik değerleri için kullanılabilir.		
C)Anlaşılabilir kurallar oluşturulabilir.		
D)Küçük ağaçları yorumlamak kolaydır.		
E)Karar ağacı oluşturmak zahmetsizdir.		
Soru10.4 Aşağıdakilerden hangisi karar ağacı algoritmalarından olan ID3'ün enropy formülüdür?(pi:olasılık sayısıdır)		
A) Σ(pi log(1/pi))		
B) ∏(pi log(1/pi))		
C) Σ(pi log(pi))		
D) ∏(log(1/pi))		
E) Hiçbiri		

Soru10.5 Karar ağaçlarında kök düğüm nasıl belirlenir?(CEVAP YOK)

A)Kazanç değeri en küçük olandır.

B)Kazanç değeri en büyük olandır.

C)Kazanç değerine bakılmaz.

D)Kazanç değeri ortanca olandır.

E)Hiçbiri.

-----

### Metin Madenciliği

Soru12.1 - ..... İlginç ve önemsiz olmayan örüntüleri veya bilgiyi çıkarma amacı için ,biçimsiz ve sayıca çok doküman analiz etme teknolojisidir.

- a) Metin Süzme
- b) Metin Madenciliği
- c)Metin Kategorizasyonu
- d) Veri Madenciliği
- e)Hiçbiri

Soru12.2 - Hangisi Metin Kategorizasyonun uygulama alanlarından biridir?

- a) Dokümanlara kategori ataması yapmak
- b) Belli bir metin içersinde sınırlı şekillerde bilgi edinimini sağlamak
- c) İlgisiz metinleri ve metin parçalarını filtrelemek
- d) Metinleri, kategori bazlı işleme mekanizmalarına yönlendirmek
- e) Hepsi

Soru12.3- Ön işleme genel adımlarının doğru sıralanışı nasıl olmadır?

- a) Kategoriler belirlenir-Kelimeler teker teker incelenir-Joker kelimeler seçilir- Vektör ağırlıkları hesaplanır
- b) Vektörün ağırlıkları hesaplanır- Joker kelimeler seçilir- Kategoriler belirlenir- Kelimeler teker teker incelenir
- c) Kelimeler teker incelenir- Vektörün ağırlıkları hesaplanır- Joker kelimeler seçilir-Kategoriler belirlenir
- d) Joker kelimeler seçilir- Kategoriler belirlenir- Kelimeler teker teker incelenir-Vektör Ağırlıkları hesaplanır
- e) Hiçbiri

Soru12.4 – Joker yöntemine göre aşağıdaki kelimelerden hangisi için ayrı sınıflandırma gerekir?

- a) Araba
- b) Arabalar
- c) Arabayı
- d) Arabacı
- e) Arabadan

Soru12.5 –Aşağıdaki sözlüğe göre "<u>Taraftarlar İstanbul'da</u> <u>maça alınmadı.</u>"metnin vektör ağırlıklandırılması nasıl olmalıdı?

```
(Taraftar* İstanbul* Gol* Futbolcu* Maç*)
a) (0,0,0,0,0)
b)(1,1,1,1,1,)
c)(0,0,1,0,0)
d)(1,1,0,1,1)
e)(1,1,0,0,1)
gini algoritması
Soru13.1 Gini algoritmasının çalışma mantığı nedir?
  a) Maksimum ağaç oluşturmak
  b) Ağacın derinliğini belirlemek
  c) Test verilerini ağaca uygulamak
  d)Hepsi
  e)Test verilerini elde etmek
 Soru13.2 Gini algoritmasında düğümler kaça ayrılır?
 a)1
 b)2
 c)3
 d)4
 e)5
Soru13.3 Gini algoritması için aşağıdakilerden hangisi doğrudur? (Etkili hesaplama için:herbir
öznitelik)
  a)Değerler üzerinde öznitelikler sıralanır.
  b)Bu değerler doğrusal olarak taranır.
  c)Her seferinde sayı matrisi güncellenir ve gini indexi hesaplanır.
  d)En küçük gini indexine sahip olan bölünme pozisyonu seçilir.
  e)Hepsi
Soru13.4 Gini algoritmasının 3 temel adımı hangisinde doğru olarak verilmiştir?
1) Maksimum ağaç oluşturmak
2) Ağacın derinliğini belirlemek
```

3) Test verilerini ağaca uygulamak

A)1-2-3 B)3-2-1 C)2-3-1 D)2-1-3 E) 1-3-2

Twoing Algoritması		
Soru15. 1) Aşağıdakilerden hangisi Twoing Algoritması'nda kullanılan parametrelerden değildir?		
a) Q(s/t)		
b) P <sub>sol</sub>		
c) P(j/t <sub>sağ</sub> )		
<del>d)</del> H(R)		
Soru15.2) Aşağıdakilerden hangisi Karar Ağaçları ile Sınıflandırma yönteminde kullanılan algoritmalardan biridir?		
<del>a)</del> Twoing Algoritması		
b) K-En Yakın Komşu Algoritması		
c) Basit Genetik Algoritma		
d) Geriye Yayma Algoritması		
Soru15.3) Twoing algoritmasında ilk olarak eğitim kümesi ikiye ayrılırken aşağıdaki parametlerdel hangi ikisini kullanır?		
a) P <sub>sol</sub> ve P <sub>sağ</sub>		
b) $P(j/t_{sol})$ ve $P(j/t_{sag})$		
e) t <sub>sol</sub> ve t <sub>sağ</sub>		
d) Gini <sub>sol</sub> ve Gini <sub>sağ</sub>		
K-MEANS		
Soru16. 1-) K-means (K-ortalama) yöntemi kaç aşamadan meydana gelir?		
a) 2 aşamadan b) 3 aşamadan c) 4 aşamadan d) 5 aşamadan e) Hiçbiri		

Soru16.2-) Aşağıdakilerden hangisi K-Means kümeleme yönteminin avantajlarındandır?

- a) Küme sayısının okunan bir parametre olması analizi esnek hale getirmektedir.
  b) Algoritmanın uygulanması kolaydır ve hızlı çalışmaktadır.
  c) Değişik dağılımlarda başarılı sonuçlar alınabilmektedir.

- d) Kategorik verilerle çalışacak şekilde adapte edilebilmektedir. e) Hepsi Soru16.3-) Aşağıdakilerden hangisi K-Means kümeleme yönteminin dezavantajlarındandır?
  - - a) Çakışan kümelerde iyi sonuç vermez.
    - b) Her eleman aynı anda verilen bir kümenin içindedir veya dışındadır.
    - c) K-means algoritması sadece sayısal veriler ile kullanılabilmektedir.
    - d) Kategorik verilerin kümelenmesi için k-means algoritması bir çözüm sunmaz
    - e) Hepsi
  - Soru16.4-) Aşağıda K-means algoritmasının adımları verilmiştir.
    - I-) Veri kümesinden k adet örnek başlangıç küme merkezleri olarak rastgele seçilir.
- II-) Öklid mesafesi kullanılarak kalan örneklerin en yakın olduğu küme merkezleri belirlenir.
  - III-) Belirlenecek küme sayısı k seçilir.
  - IV-) Her küme için yeni örneklerle küme merkezleri hesaplanır.

Bu adımların doğru sıralanışı aşağıdakilerden hangisidir?

- a) IV -I- II-III
- b) III- II-IV-I
- c) III- I-II-IV
- d) IV-II-III-I

Soru16.5-) K- means kümelemede k küme sayısı başlangıçta nasıl belirlenir?

- a) Tahmini bir sayı verilir.
- b) En küçük öklid uzaklığına eşittir.
- c) Veri sayısının yarısı değerindedir.
- d) Algoritma çalışırken doğru sayıyı bulur.
- e) Hiçbiri.

web madenciliği

Soru17.1 Aşağıdakilerden hangisi web yapı madenciliğinin görevlerindendir?

- a) web kaynaklarının içeriklerinden yararlı bilgiyi elde etmek
- b) web sayfaları arasındaki bağlantıları incelemek
- c) web sayfalarının kullanma durumlarını incelemek
- d) web üzerindeki çeşitli sunucularda tutulan kullanıcı erişim hareketlerinden veri elde etmek
- e) web sayfasına yapılan kullanıcı erişim kayıtlarının analiz edilmesi

Soru17.2 Aşağıdakilerden hangisi web kullanım madenciliği verilerinden biri değildir?

- a) Veri Yapısı
- b) Veri İçeriği
- c) Kullanım Verileri
- d) Kullanıcı Profili Verileri

# e) Site yapısı ve içeriği

Soru17.3 Aşağıdakilerden hangisi web madenciliği veri kaynaklarından biri değildir?

- a) Şirket web sitelerib) Arama motorları
- c) Web sayfaları arasındaki bağlantılar.
- d) Sosyal ağlar
- e) Tartişma grupları

Soru17.4 Aşağıdakilerden hangisi web yapı madenciliğinin ürettiği bilgilerden biri değildir?

- a) Sayfanın Kalitesi
- b) Web Sayfalarının Sınıflandırılması
- c) İlişkili Sitelerin (sayfaların) Bulunması
- d) Siteye yapılan erişim kayıtları
- e) Tekrarların Bulunması

\_\_\_\_\_\_

Tüm Yönleriyle Veri Madenciliği ve Geleceği

Soru18.1)Aşağıdakilerden hangisi VTBK sürecinde izlenmesi gereken temel aşamalardan biri değildir?

A)Problemin tanımlanması

B)Verilerin hazırlanması

C)Modelin kullanılması

D)Modelin sonuçlandırılması

Soru18.2) Aşağıdakilerden hangisi Veri Madenciliğinde Karşılaşılan Zorluklar'dan değildir?

B) İlgisiz A) Eksik veri veri

C) Dinamik veri

D) Artık veri

Soru18.3) Aşağıdakilerden hangisi veri madenciliğinde Kümeleme Teknikleri arasında yer almaz?

- A) Bölme Yöntemleri
- B) Hiyerarşik Yöntemler
- C) Izgara Tabanlı Yöntemler
- D) Toplama Yöntemleri

Soru18.4)Aşağıdakilerden hangisi VTBK sürecinde kullanılan Veri Kaynakları'ndan değildir?

A)Eski Saklanan Veriler

B)ERP Veritabanı

C)Sürdürülebilir Veriler

D)Dış Kaynaklı veriler

Soru18.5)Aşağıdakilerden hangisi Veri Madenciliği kullanım alanlarından biri değildir?

# A)Tahta Üretim Sektörü

B)Sağlık Sektörü

C) Finans Sektörü

D)Telekominikasyon Sektörü

\_\_\_\_\_

aprori

Soru 19.1-2-3:

TID	Item_ID Listesi
T100	I1, I2, I5
T200	I2, I4
T300	12, 13
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Tablo 4.1 İşlem verileri

Tablo 4,1'de bir marketten yapılan alışveriş bilgilerini içeren veri tabanı görülmektedir. Bu veri tabanında yapılan alışverişlerin numaraları TID sütununda görülmektedir. Her alışverişte satın alınan ürünler de Item ID Listesi sütununda görülmektedir. Bu tabloya göre aşağıdaki soruları cevaplayınız?

Soru19.1 Hangi 2 ürün hiç beraber alınmamıştır?

- A) I1-I5
- B) 12-I4
- C) I1-I3
- D) I4-I5
- E) 13-15

Soru19.2 Bu markette aşağıdakilerden hangi iki ürünü rafta yan yana koymak daha mantıklıdır?

- A) I2-I5
- B) I1-I4
- C) I2-I3
- D) I4-I5
- E) I2-I5

Soru19.3 X ürünü için destek, tüm alışverişler içinde X ürününün oranıdır. Buna göre I2 ürünü için bu oranı aşağıdaki destek formülünden yararlanarak bulunuz?

# $Destek(X) = \frac{X \text{ sayısı}}{Toplam \text{ Alışveriş sayısı}}$

A)	3/9
	ر رد

B) 4/9

C) 5/9

D) 7/9 E) 8/9

Gerçek Veriler ile Veri Ambarı Uygulama

Soru20.1-) Veri ambarları için çok boyutlu perspektifi gerçekleyebilecek kaç çeşit yeni modelleme tekniği keşfedilmiştir?

- A -) 2
- B -) 3
- C -) 4
- D-) 5
- E-) 1

Soru20.2-) Aşağıdakilerden hangisi Veri Ambarı Yapısı içinde bulunmaz?

- A-) Stagging Area
- B-) External Data
- C-) Data Marts
- D-) Data Vault
- E -) Hiçbiri

Soru20.3-) External Data Nedir?

- A-) Verilerin geçici olarak saklandığı yerdir.
- B-) Organizasyonun (işletmenin) belirli kullanıcıları için ayrılmış ve onlara ait verileri içerir.

C-) Harici kaynaktan gelen veri çeşitleri; xml dosyları, txt, cvs, mesaj sistemleri olacağı gibi, çok farklı veri tiplerini kapsar.

- D-) Veri Ambarındaki elementler arasında ki ilişkiyi belirler.
- E-) Hiçbiri

Soru20.4-) Aşağıdakilerden hangisi Veri Ambarının Temel Özelliklerinden Değildir?

A-) Subject Oriented(Konu Yönelimlilik)

- B-) Integrated(Bütünleşmiş)
- C-) Time-variant(Zaman-Değişimlilik)
- D-) Non-Volatility(Kalıcılık)
- E-) Hiçbiri

Soru20.5-) Aşağıdakilerden hangisi Veri Ambarı İçinde Yer Almaz?

- A-) Güncel Veriler
- B-) Sayısal Veriler
- C-) Düzeltilmiş Veriler
- D-) Eski Veriler
- E-) Birleştirilmiş Veriler