

**BAŞKENT ÜNİVERSİTESİ**  
TİCARİ BİLİMLER FAKÜLTESİ  
Yönetim Bilişim Sistemleri Bölümü  
TBF 491

## VERİ MADENCİLİĞİ

Dr. Ali Serhan KOYUNCUGİL

**BAŞKENT ÜNİVERSİTESİ**  
TİCARİ BİLİMLER FAKÜLTESİ  
Yönetim Bilişim Sistemleri Bölümü  
TBF 491

## BÖLÜM 6 VERİ MADENCİLİĞİ METODOLOJİSİ

Dr. Ali Serhan KOYUNCUGİL

## BÖLÜM 6 VERİ MADENCİLİĞİ METODOLOJİSİ

**6.1 Veri Madenciliği İçin Sanayilerarası Standart Süreç (VM-3S/CRISP-DM)**

**6.2 SEMMA**

**6.3. Diğer Veri Madenciliği Metodolojileri**

**6.4. Veri Madenciliği Metodolojilerinin Genel Değerlendirmesi**

## BÖLÜM 6 VERİ MADENCİLİĞİ METODOLOJİSİ

Veri madenciliği için hangi temel metodolojiyi kullanıyorsunuz?  
(KDNugget, Nisan 2004)

CRISP-DM (72)	42%
SEMMA (17)	10%
Kendi organizasyonumun (11)	6%
Kendi metodolojim (48)	28%
Diğer (10)	6%
Yok (12)	7%

### 6.1 VERİ MADENCİLİĞİ İÇİN SANAYİLERARASI STANDART SÜREÇ (VM-3S/CRISP-DM)

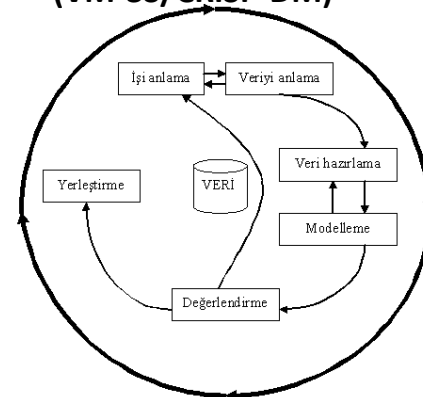
### 6.1 VERİ MADENCİLİĞİ İÇİN SANAYİLERARASI STANDART SÜREÇ (VM-3S/CRISP-DM)

- Daimler-Benz AG (Almanya),
  - SPSS/Integral Solutions Ltd. (İngiltere)
  - NCR Systems Engineering Copenhagen (Danimarka)
- aralarına sonradan katılan
- OHRA Verzekeringen en Bank Groep B.V. (Danimarka)
- ile birlikte veri madenciliğini halihazırda uygulayan ve kullanan firmalar olarak kendi başlarına uyguladıkları süreçlerin doğruluğunu ve bir standardın oluşturulup oluşturulamayacağını sorgulamışlardır.

### 6.1 VERİ MADENCİLİĞİ İÇİN SANAYİLERARASI STANDART SÜREÇ (VM-3S/CRISP-DM)

- Tüm uygulayıcılara yönelik, veri madenciliği için iyi düşünülmüş, kimseye ait olmayan, ücretsiz bir standart süreç modeli hazırlamak üzere yola çıkmışlardır.
- 1999 yılı sonunda Veri Madenciliği için Sanayilerarası Standart Süreç (VMSSS) (CRoss-Industry Standard Process for Data Mining- CRISP-DM) CRISP-DM 1.0 versiyonunu hazırlamış ve veri madenciliği için standart bir süreç modeli önermişlerdir.

### 6.1 VERİ MADENCİLİĞİ İÇİN SANAYİLERARASI STANDART SÜREÇ (VM-3S/CRISP-DM)



## 6.2 SEMMA

## 6.2 SEMMA

SEMMA, veri madenciliğinde yürütülen sürecin özünü gösteren;

- Örnekle (**S**ample),
- Araştır (**E**xplore),
- Değiştir (**M**odify),
- Modelle (**M**odel),
- Değerlendir (**A**ssess)

sürecinin başharflerinden oluşmaktadır.

## 6.3. DİĞER VERİ MADENCİLİĞİ METODOLOJİLERİ

## 6.3. DİĞER VERİ MADENCİLİĞİ METODOLOJİLERİ

### CABENA VD'nin (1997) METODOLOJİSİ

Genel olarak veri madenciliği adımları;

1. Adım: İş amaçlarının tanımlanması
2. Adım: Veri hazırlanması
3. Adım: Veri dönüşümü
4. Adım: Veri madenciliği
5. Adım: Sonuçların analizi
6. Adım: Bilginin özümsemesi

biçimindedir.

### **6.3. DİĞER VERİ MADENCİLİĞİ METODOLOJİLERİ**

#### **CHEN'in (2001) METODOLOJİSİ**

1. Adım: Problemi tanımla
2. Adım: Veri biriktir/seç
3. Adım: Veri hazırla
4. Adım: Veri ön-işlemesi
- 5a. Adım: Model veya algoritma seç
- 5b. Adım: Model veya algoritmanın eğitim parametrelerini seç
6. Adım: Verinin eğitimi/testi veya algoritmanın uygulanması
7. Adım: Son değerlendirme/modelin entegrasyonu

### **6.3. DİĞER VERİ MADENCİLİĞİ METODOLOJİLERİ**

#### **TWOCROWS'un (1999) METODOLOJİSİ**

1. Adım: İş problemini tanımla
2. Adım: Veri madenciliği veri tabanını kur
3. Adım: Veriyi keşfet
4. Adım: Modelleme için veriyi hazırla
5. Adım: Modeli kur
6. Adım: Modeli değerlendir
7. Adım: Modeli ve sonuçları yerleştir

biçiminde yer almaktadır.

### **6.4. VERİ MADENCİLİĞİ METODOLOJİLERİNİN GENEL DEĞERLENDİRMESİ**

### **6.4. VERİ MADENCİLİĞİ METODOLOJİLERİNİN GENEL DEĞERLENDİRMESİ**

Tüm veri madenciliği metodolojileri dikkate alındığında, VTBK iş akışının, hemen hemen hepsinde yer aldığı görülmektedir:

- Hedef veri seçimi
- Veri Hazırlama / Veri Ön işleme / Veri Temizleme
- Veri dönüşümü
- Veri madenciliği
- Yorumlama
- Raporlama