

**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 1 VERİ MADENCİLİĞİNE GİRİŞ

Dr. Ali Serhan KOYUNCUGİL

### BÖLÜM 1 VERİ MADENCİLİĞİNE GİRİŞ

- 1.1. Veri Madenciliği Ve Veri Tabanlarında Bilgi Keşfi
- 1.2. Veri Ambarı
- 1.3. Veri Madenciliğinin Tanımı
- 1.4. Veri Madenciliğinin Tarihçesi
- 1.5. Veri Madenciliğinin Uygulama Alanları
- 1.6. İstatistiksel Öğrenme

### 1.1 VERİ MADENCİLİĞİ VE VERİ TABANLARINDA BİLGİ KEŞFİ

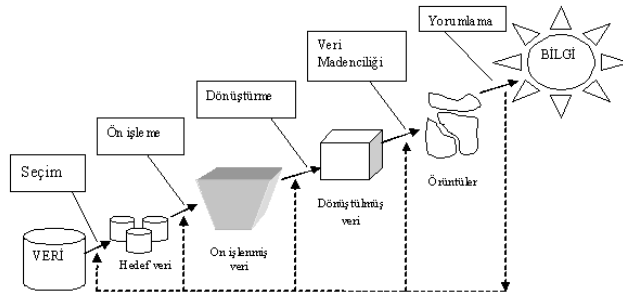
### 1.1.1 Veri Tabanlarında Bilgi Keşfinin Tanımı

Veri Tabanlarında Bilgi Keşfi, veriden faydalı bilginin keşfedilmesi sürecinin tamamına atıfta bulunmakta ve veri madenciliği bu sürecin bir adımına karşılık gelmektedir. Veri madenciliği, veriden örüntülerin aktarımı için özel algoritmaların uygulanmasıdır.

### 1.1.1 Veri Tabanlarında Bilgi Keşfinin Tanımı

VTBK, verinin nasıl depolanıp erişileceğinden, algoritmaların devasa veri setlerine nasıl ölçeklenebileceğine ve hala etkin olarak çalışmalarına, sonuçların nasıl yorumlanabileceği ve görselleştirilebileceğine ve bütün insan-makine interaksyonunun kullanışlı olarak nasıl modellenip, desteklenebileceğine olmak üzere veriden bilginin keşfinin tüm süreçleri üzerine odaklanır.

### 1.1.2. Veri Tabanlarında Bilgi Keşfi Süreci



Şekil 1. VTBK Sürecinin Adımları

### 1.1.2. Veri Tabanlarında Bilgi Keşfi Süreci

- 1.Adım: Uygulama alanı ve ilgili önsel bilgi ile ilgili bir anlayış geliştirmek ve müşterinin bakış açısından VTBK sürecinin hedefini tanımlamak.
- 2.Adım: Hedef veri kümesini yaratmak: Keşfin uygulanacağı veri kümesini seçmek veya değişkenlerin bir alt kümesi veya veri örnekleri üzerine odaklanmaktır.

### **1.1.2. Veri Tabanlarında Bilgi Keşfi Süreci**

*3.Adım: Veri temizleme ve ön işleme: Eğer uygunsa gürültünün kaldırılması, model için gerekli enformasyonun toplanması, kayıp veri alanları için stratejilere karar vermeyi içeren temel operasyonlardır.*

*4.Adım: Veri indirgeme ve projeksiyon: Görev hedefine bağlı veriyi temsil edecek faydalı özellikleri bulmaktır. Boyut indirgeme veya dönüşüm yöntemleriyle göz önüne alınan değişken sayısı indirgenebilir veya verinin değişmez (invariant) temsili bulunabilir.*

### **1.1.2. Veri Tabanlarında Bilgi Keşfi Süreci**

*5.Adım: VTBK sürecinin hedefleri ile (1. Adımdaki), veri madenciliği görevinin eşleştirilmesi: Karakterizasyon (Betimleme), Kümeleme, Sınıflandırma, Kural ve Ağaç Çıkarımı, Modelleme ve Birlikte görevlerinden bir veya birden fazlası seçilecektir.*

*6.Adım: Veri madenciliği algoritma(larının)sının seçimi: Açıklayıcı analizler, model ve hipotez seçimi: Tercih edilen vm algoritmaları ve seçilen yöntemler veri örüntülerini araştırmak için kullanılır. Bu süreç, hangi modelin ve parametrelerin uygun olabileceğine ve vm yönteminin VTBK sürecinin bütün kriterleriyle eşleşip eşleşmediğine karar verilmesini içermektedir.*

### **1.1.2. Veri Tabanlarında Bilgi Keşfi Süreci**

*7.Adım: Veri Madenciliği: Özel bir temsili form veya temsili küme içerisinde ilgilenilen örüntüler; sınıflandırma yı, kümelemeyi, kuralları ve ağaçları, modellemeyi ve ilişkileri içererek biçimde araştırılır.*

*8.Adım: Elde edilen örüntülerin yorumlanması: Sonraki iterasyonlarda, Adım 1-7'den herhangi birine dönüşmesi ihtimaliyle elde edilen örüntüler yorumlanır.*

### **1.1.2. Veri Tabanlarında Bilgi Keşfi Süreci**

*9.Adım: Keşfedilen bilgilerin birleştirilmesi: Keşfedilen bilgi sonraki çalışmalar için bir başka sistem altında toplanabilir veya basitçe dokümanite edilip, raporlanarak ilgili birimlere iletilir. Bu aynı zamanda, önceden inanılan veya aktarılan bilgilerin doğruluğunu kontrol etme ve olası farklılıkların ayrıştırılmasını da içerir.*

**1.1.3. Veri Tabanlarında Bilgi Keşfi Sürecinin  
Veri Madenciliği  
Adımı**

Bilgi keşfi hedefleri, sistemin kullanım amacına göre tanımlanır. Hedefleri ikiye ayırabiliriz:

1. Doğrulama
2. Keşif

**1.1.3. Veri Tabanlarında Bilgi Keşfi Sürecinin  
Veri Madenciliği  
Adımı**

Doğrulama ile, sistem kullanıcının hipotezlerini doğrulamak ile sınırlıdır.

Keşif ile, sistem bağımsız olarak yeni örüntüler bulur.

Keşif hedefi,

- bazı varlıkların gelecekteki davranışlarını tahmin etmek için sistemin örüntüleri bulmasında kullanıldığında tahmin
- kullanıcıya sunumda insanın anlayabileceği bir form için sistem kullanıldığında tanımlama olmak üzere iki alt gruba ayrılmaktadır.

**1.1.3. Veri Tabanlarında Bilgi Keşfi Sürecinin  
Veri Madenciliği  
Adımı**

Veri madenciliği gözlenen veriye model uydurmayı veya gözlenen verideki örüntüleri tanımlamayı gerektirmektedir.

Model uydurma, bilgi çıkarımı rolünü üstlenmektedir: Modelin, kullanışlı veya ilginç keşifsel bilgiye işaret edip etmediği, tamamıyla interaktif VTBK sürecinin subjektif insan yargısına tipik olarak ihtiyaç duyduğu bir parçasıdır.

**1.1.3. Veri Tabanlarında Bilgi Keşfi Sürecinin  
Veri Madenciliği  
Adımı**

Model uydurmada istatistiksel ve mantıksal olmak üzere iki temel matematiksel yapı kullanılmaktadır:

Modelde,

- istatistiksel yaklaşım deterministik olmayan etkiye,
- mantıksal yaklaşım ise deterministik etkiye izin vermektedir.

## 1.2. VERİ AMBARI

### 1.2.1. VERİ AMBARININ TANIMI

Veri Ambarları, Veri Madenciliği ile eşanlı olarak anılan ve Veri Madenciliği sürecinin gerçekleştirildiği veriyi sağlayan özel bir veri tabanıdır. Tanım olarak Veri Ambarı, pekçok farklı kaynaktan ve genellikle de farklı yapıda verinin depolandığı ve hepsinin de aynı birleşik çatı altında kullanılmasının ümit edildiği yapılardır.

Ayrıca, Veri Ambarı pek çok farklı kaynaktan elde edilen, farklı formatta veriyi aynı çatı altında analiz etme imkanı tanır.

Veri ambarcılığı ise veri ambarını kurma ve kullanma sürecine denmektedir.

### 1.2.2. VERİ AMBARININ ÖZELLİKLERİ

**Kurumun operasyonel veri tabanından ayrı olarak yapılandırılmış bir karar destek veritabanıdır.**

**Konsolide edilmiş, tarihi verileri sağlayarak bilgi işlemeyi destekler.**

**Bir veri ambarı, yöneticilerin karar verme sürecine destek olmak üzere verinin**

- Konuya yönelik
- Entegre
- Zamana bağlı değişen (zamana endeksli)
- Değişmez

**halini sunmaktadır.**

### 1.2.3. VERİ AMBARININ AMACI

**1. Veri temizleme**

**2. Veri erişimi**

### 1.3. VERİ MADENCİLİĞİNİN TANIMI

### 1.3. VERİ MADENCİLİĞİNİN TANIMI

Veri madenciliği, pek çok analiz aracı kullanımıyla veri içerisinde örüntü ve ilişkileri keşfederek, bunları geçerli tahminler yapmak için kullanan bir süreçtir.

Veri madenciliği, en basit tanımıyla, veri tabanlarındaki ilişkili örüntüleri **OTOMATİK** olarak belirlemedir.

Veri madenciliği veri kümesi içerisinde keşfedilmemiş örüntüleri bulmayı hedefleyen teknikler koleksiyonunu betimlemektedir. Veri madenciliğinin amacı, geçmiş faaliyetlerin analizini temel alarak gelecekteki davranışların **TAHMİNİNE** yönelik karar-verme modelleri yaratmaktır.

### 1.3. VERİ MADENCİLİĞİNİN TANIMI

Koyuncugil & Özgülbaş'a göre Veri madenciliği, (büyük hacimde) veri içerisindeki gizli tahminsel bilgi, örüntü, ilişki, benzerlik veya benzemezliğin otomatik olarak açığa çıkarılması ve aktarılması sürecinde kullanılan akıllı algoritmalarla evrimleşmiş istatistiki analiz, makine öğrenimi ve örüntü tanıma yöntemleri koleksiyonudur.

### 1.3. VERİ MADENCİLİĞİNİN TANIMI

Veri madenciliği henüz gelişmekte olan bir alan olduğu için pek çok farklı tanım yapmak mümkündür. Buna karşın,

Veri madenciliğinin tek hedefi düşük düzeyde enformasyon sağlayan veriden, yüksek düzeyde kıymetli bilgiyi açığa çıkarmaktır.

## 1.4. VERİ MADENCİLİĞİNİN TARİHÇESİ

## 1.4. VERİ MADENCİLİĞİNİN TARİHÇESİ

### Veri Madenciliğinin Tarihçesi

Veri Madenciliği'nin altyapısının oluştuğu süreç kronolojik olarak aşağıda verilmektedir.

#### 1950'ler

İlk bilgisayarlar (Sayımlar için bilgisayarlar kullanılıyor)

#### 1960'lar

Veri koleksiyonları  
Veritabanı yaratımı (Hiyerarşik ve ağ modelleri)

## 1.4. VERİ MADENCİLİĞİNİN TARİHÇESİ

### Veri Madenciliğinin Tarihçesi

#### 1970'ler

İlişkisel veri modeli  
İlişkisel VTYS uygulamaları

#### 1980'ler

İlişkisel VTYS yaygınlaşıyor  
Uygulamaya yönelik VTYS (Mekansal, Bilimsel, Mühendislik, vs.)

## 1.4. VERİ MADENCİLİĞİNİN TARİHÇESİ

### 1990'lar

Günlük işlemlerden derlenen büyük miktarda verinin nasıl değerlendirilebileceği sorgulanmaya başlıyor. Bu noktada bahsedilmesi gereken birkaç önemli olay söz konusu:

**1989**, KDD (IJCAI)-89 Veri Tabanlarında Bilgi Keşfi Çalışma Grubu toplantısı

**1991**, KDD (IJCAI)-89'un sonuç bildirgesi sayılabilecek 'Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop' makalenin KDD ile ilgili temel tanım ve kavramları ortaya koyması

## **1.4. VERİ MADENCİLİĞİNİN TARİHÇESİ**

**1992,** Veri Madenciliği konusunda ilk yazılımın geliştirilmesi

**1995,** 1. Uluslararası Bilgi Keşfi ve Veri Madenciliği Konferansı'nın (KDD-95) açılış konuşması

## **1.4. VERİ MADENCİLİĞİNİN TARİHÇESİ**

### **2000'ler**

Veri Ambarları,

Veri Madenciliği yaygınlaşıyor. (Son 10 yılda veri depolama ünitelerinin fiyatlarında sürekli bir düşüş var.)

## **1.5. VERİ MADENCİLİĞİNİN UYGULAMA ALANLARI**

## **1.5. VERİ MADENCİLİĞİNİN UYGULAMA ALANLARI**



### 1.5. VERİ MADENCİLİĞİNİN UYGULAMA ALANLARI

Büyük hacimde  
veri olan her  
yerde veri  
madenciliği  
kullanılabilir.

### 1.6. İSTATİSTİKSEL ÖĞRENME

### 1.6. İSTATİSTİKSEL ÖĞRENME

Veri Madenciliği genel anlamda incelendiğinde, sürecin işleyişi, içeriği ve bileşenleri bakımından; İstatistik ve Bilişim olmak üzere temel anlamda iki bakış açısına sahip olmak mümkündür.

Veri madenciliğine istatistiksel bakış İstatistiksel Öğrenme olarak adlandırılmaktadır. İstatistiksel Öğrenme zaman zaman Veriden Öğrenme olarak da isimlendirilebilmektedir.

İstatistiksel perspektiften veri madenciliği ise akıllı algoritmalarla evrimleşmiş istatistiksel analiz yöntemlerinin, otomatik tahmin için kullanımı olarak tanımlanabilir.