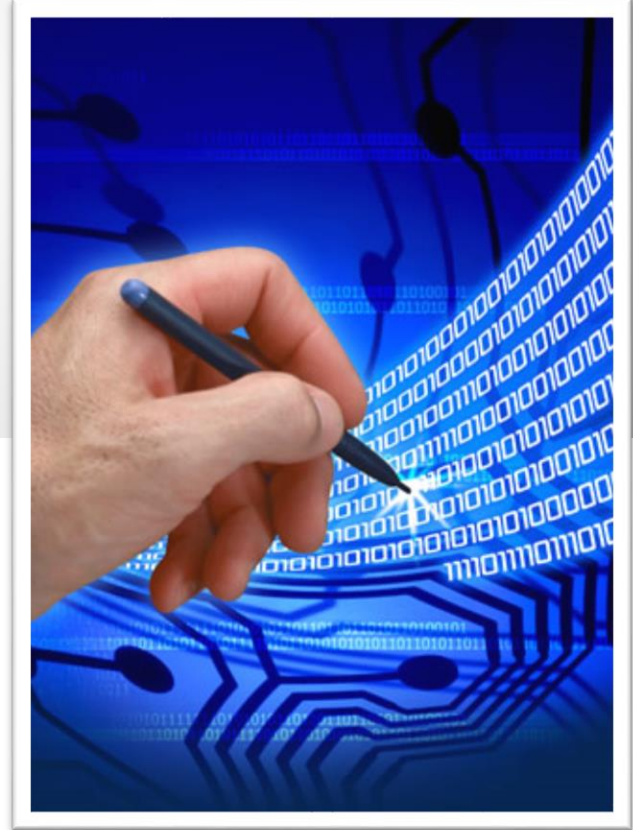


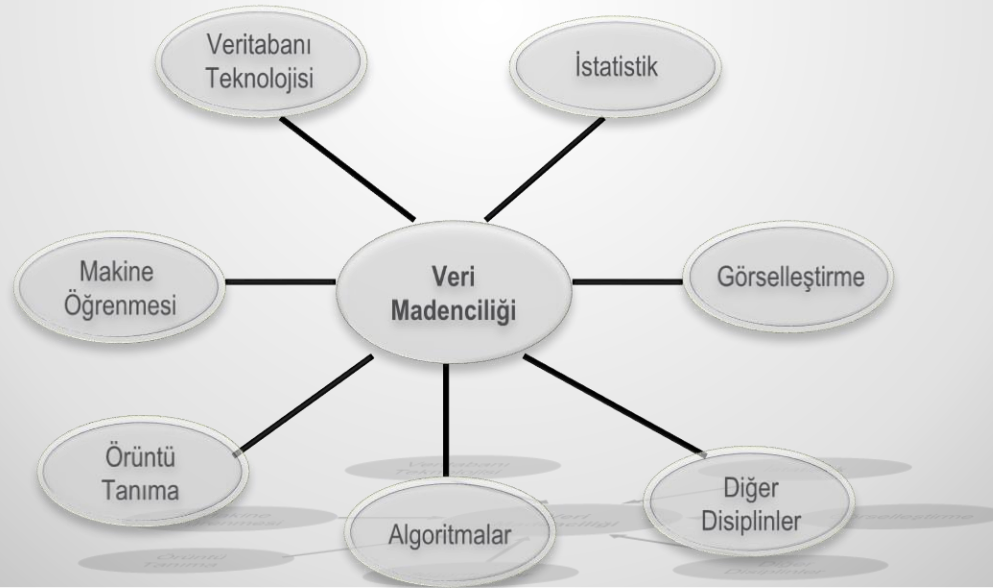
# *Veri Madenciliđi*

## Uygulama Alanları Ve Örnekler



# Veri Madenciliği Nedir :

“Verilerden üstü kapalı, çok net olmayan , önceden bilinmeyen ancak potansiyel olarak kullanılabilir bilgi ve örüntülerin çıkarılması olarak tanımlanabilir.”



# Veri Madenciliği Süreci

Problemin  
Tanımı

Kullanılacak  
verilerin  
seçilmesi ve  
hazırlanması

Verilerin  
bulunması ve  
analizi

Modelin  
oluşturulması

Modelin  
geçerliliğinin  
testi

Bilginin  
üretilmesi,  
eylem planına  
dönüştürülmesi  
ve sonuçların  
ölçülüp  
değerlendirilme  
si

## **Neden Veri Madenciliği?**

- ☐ **Mağazalardaki alış/satış işlemleri**
- ☐ **Banka ve Kredi kartı işlemleri**
- ☐ **Bilimsel veriler, uydu ve radarlardaki algılayıcılardan gelen veriler**
- ☐ **Web teknolojilerinden gelen veriler**
- ☐ **Bir çok sektördeki veri ve işlemler**

Veri madenciliğinin karar verici için;

**Mevcut müşterilerin karar verici tarafından daha iyi tanınmasını sağlayabilir.**

- **Bankacılık**

- Kredi taleplerinin değerlendirilmesi.
- Farklı finansal göstergeler arasında gizli korelasyonların bulunması,
- Kredi kartı üzerinden dolandırıcılıkların tespiti,
- Kredi kartı harcamalarına göre müşteri gruplarının ortaya çıkarılması,

- **Pazarlama**

- Satış tahminleri
- Müşterilerin satın alma örüntülerinin ortaya çıkarılması,
- Pazar sepeti analizi,
- Müşterilerin demografik özellikleri arasındaki bağlantıların bulunması,
- Düzenlenen kampanyalarda cevap verme oranının artırılması,
- Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması,
- Müşteri ilişkileri yönetimi ve değerlendirmesi

- **Sigortacılık**

- Dolandırıcılıklarının tespiti,
- Yeni poliçe talep edecek müşterilerin tahmin edilmesi,
- Riskli müşteri örüntülerinin ortaya çıkarılması.

# Web Madenciliği Nedir ?



Web madenciliği, veri madenciliği tekniklerinin kullanılarak web belgelerinden ve servislerinden otomatik olarak bilginin ayıklanması, ortaya çıkarılması ve tahlil edilmesidir.[Etzoni]

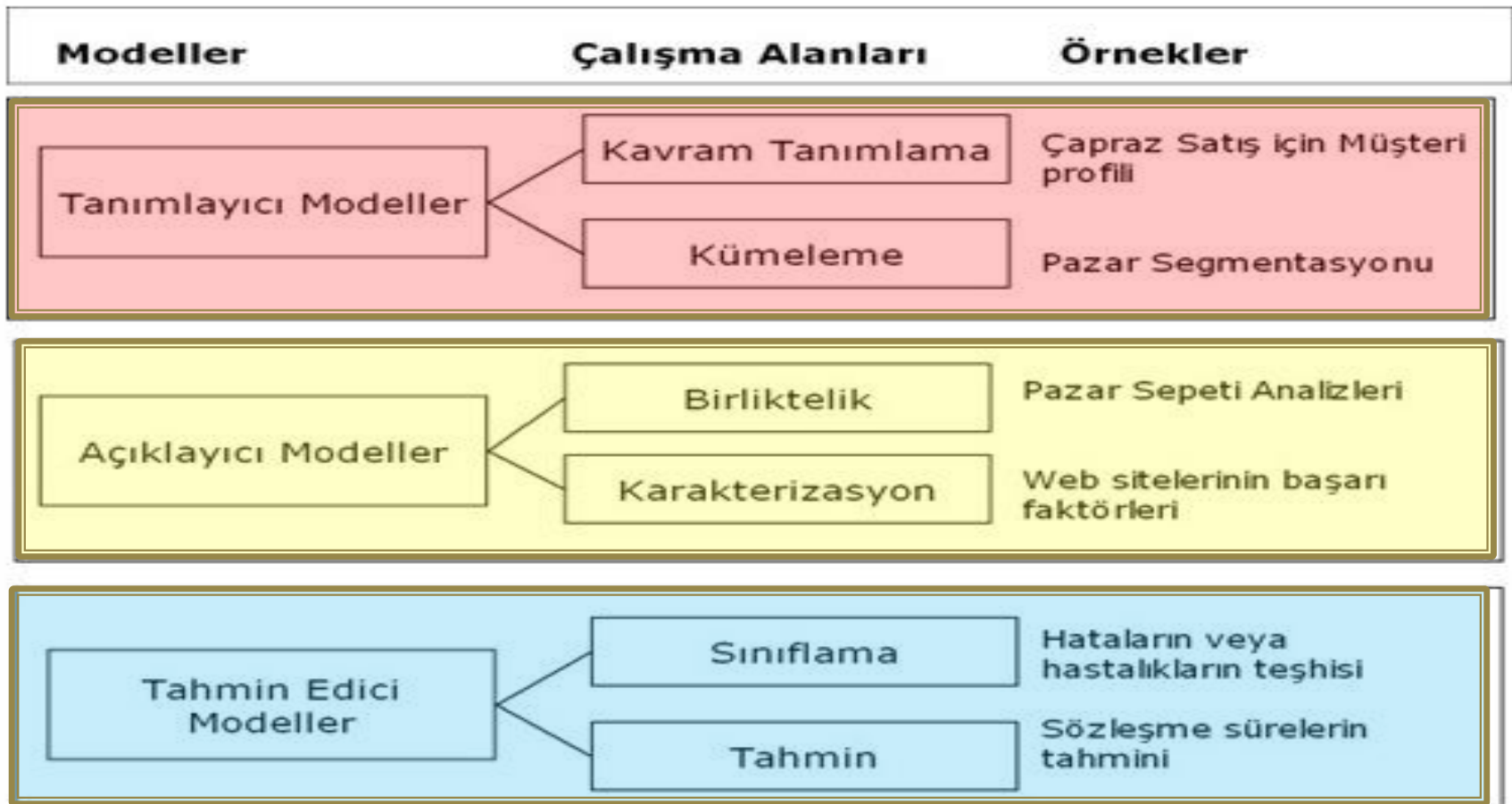
- ❖ 1 Milyarın Üzerinde Sayfa Sayısı ( 15 TeraBytes!!)
- ❖ Dinamik Yapısı
- ❖ Her Gün Yarım Milyon Yeni Sayfa
- ❖ Her Gün *Yüz Milyonlarca Sorgu*

- Kaynak Bulma (Resource finding)
- Bilgi seçimi ve ön işleme (Information selection and pre-processing)
- Genelleme (Generalization)
- Analiz (Analysis)

# E-Ticaret ve Web Madenciliği ile beklentiler ;

- ✓ Müşterilerin yaşam sürelerini öğrenmek
- ✓ Yeni stratejiler belirlemek
- ✓ Promosyonları değerlendirmek
- ✓ Hedef Kitle Seçmek
- ✓ Müşterileri tanımak

# Veri Madenciliği uygulamaları, model ve çalışma alanlarına göre düzenlenebilir.



**Veri Madenciliği, Çalışma Alanları ve Örnekler**



## Çalışma Alanları

## Bilginin Temsili

## Sunum Şekli



## Veri Madenciliği Çalışma Alanları, Bilgi Temsili ve Sunum Yapıları

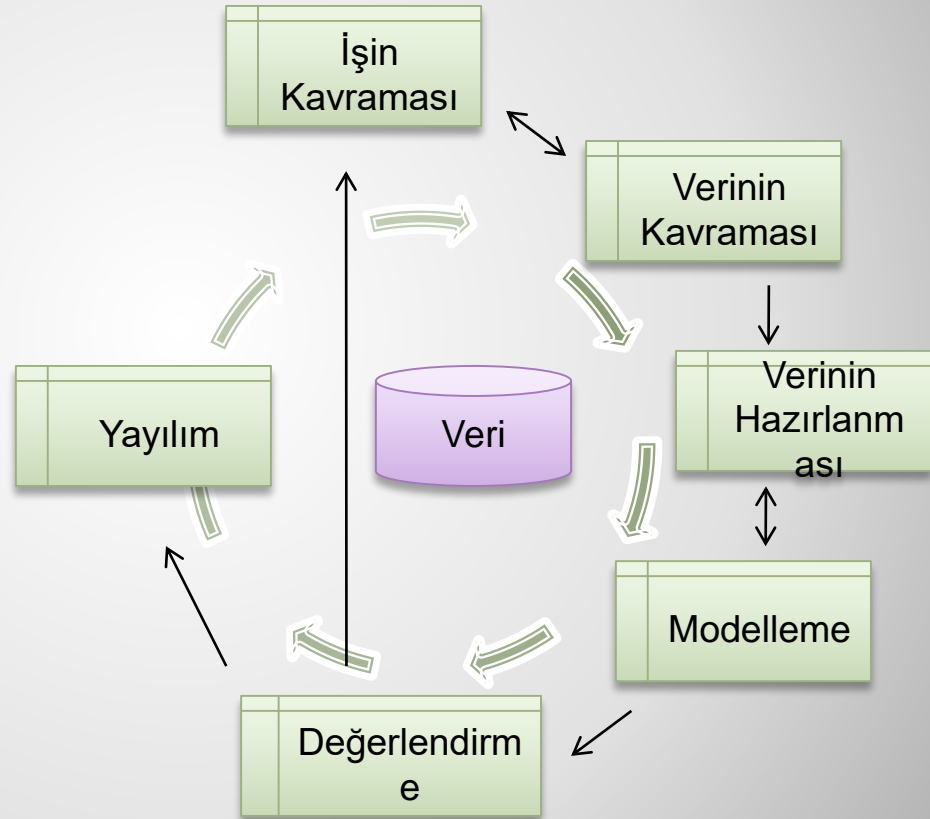
Veri madenciliği için çapraz endüstri standart süreci (CRISP-DM)

VM çalışmalarının başarılı bir şekilde sonuçlanabilmesi için sistematik bir yaklaşım zorunludur. VM araçları sunan birçok satıcı, danışma firması ve araştırmacı kullanıcılara yönelik bir çok süreç modelleri oluşturmuşlardır.

Bunlardan en yaygın olarak kullanılanı, VM araçlarını satan firmalardan bazılarının ve VM uygulamalarını işletme faaliyetlerine uyarlayan ilk kuruluşların oluşturduğu bir konsorsiyum tarafından geliştirilen *Çapraz Endüstri Veri Madenciliği Standart Süreci*'dir.

VM projesi yaşam döngüsü altı aşamayı içermektedir. Bu aşamalar aşağıda verilmiştir:

- 1.İşin Kavranması
- 2.Verinin Kavranması
- 3.Verinin Hazırlanması
- 4.Modelleme
- 5.Planı Göre Yayılma
- 6.Değerlendirme



# Sağlık Sektöründe Veri Madenciliği

- **Veri madenciliği ve Bilgi Keşfi tanımı:**
  - Verilerde daha önceden bilinmeyen, anlamlı ve değerli bilgiler elde etme işlemleridir.
- **Uzman sistemler (1970 ler), yaygınlaşamama nedenleri:**
  - Verilerin hızlı değişmesi
  - Uzmanlar arasındaki görüş ayrılıkları
- **Yapay Sinir Ağları(1990 lar)**
  - Hastaların gelecekteki sağlık durumlarının tahmini
  - Maliyet tahminleri



# Sağlık Sektöründe Veri Madenciliği Uygulamaları

- Belirli bir hastalığa sahip hastaların ortak özelliklerinin tahmin edilmesi
- Tıbbi tedavilerinin sonuçlarının tahmin edilmesi
- Hastane maliyetlerinin tahmin edilmesi
- Ölüm oranları ve salgın hastalıkların tahmin edilmesi
- Hastalıkların Yönetimi
- Hastaların sürekli doktor değiştirmelerinin nedenleri
- Hastaların hastanelerdeki kalış süreleri araştırılmış
- Hile Tespiti Çalışmaları
- İlaçların yan etkileri

# Tıp ve Biyoinformatik Alanlarında Veri Madenciliği Çalışmaları

- Tıp ve sağlık alanındaki yapısal olmayan dokümanlar için metin madenciliği çalışmaları
  - Tanı, tedavi bilgileri
  - Klinik dökümanlar
  - Bilimsel makaleler



# Tıp ve Biyoinformatik Alanlarında Veri Madenciliği Çalışmaları

- Avrupa Biyoinformatik Enstitüsü metin madenciliği araştırma grubu bilimsel makalelerden bilgi keşfi yapmak için araştırmalar yapmaktadır.
- Birçok yazılım sistemi ve aracı geliştirmiştir:
  - EBIMEd
  - Whatizit

# Tıp ve Biyoinformatik Alanlarında Veri Madenciliği Çalışmaları

Biyolojik araştırmalar tıbbi gelişmelere önemli katkılar sağlar.

- BIOMART
  - İlişkisel veriler için geliştirilen veri madenciliği sistemidir.
- OMIM(Online Mendelian Inheritance in Man)
  - National Center for Biotechnology Information(NCBI) tarafından geliştirilmiştir
  - Genetik bozukluklarla ilgili bilinen hastalıkların saklandığı veritabanıdır.



# Tıp ve Biyoinformatik Alanlarında Veri Madenciliği Çalışmaları

- İnsan Genom projesinin tamamlanması ile birlikte genom bilgileri ile klinik bilgiler arasında bağ kurabilmek için projeler ve veritabanları geliştirildi:
  - HapMap Projesi
  - DECIPHER Projesi



# Bankacılık ve Finans Sistemlerinde Veri Madenciliği

Pek çok banka ve finansal kuruluş hesap açtırma, yatırım yapma gibi banka hizmetleri, otomobil, iş, ev gibi kredi hizmetleri sunmaktadır. Bankacılıkta ve finansal sanayide kullanılan verinin tam, güvenilir ve yüksek kalitede olması sistematik veri analizi uygulanmasını kolaylaştırır.

Bankacılık sektöründe, veri madenciliği kredi sahtekarlığı tahminleme, risk değerlendirme, eylem analiz değerlendirme, kar analizi gibi alanlarda kullanılır.

Finansal kuruluşlarda, stok fiyat tahminleme, gümrük ölçümleme, portföy yönetimi, mal fiyat tahminleme gibi alanlarda yapay sinir ağları kullanılır.



# Hizmet Sektöründe Veri Madenciliği

Perakende sanayi, satış bilgileri, müşteri profili, ürün nakil bilgileri, servis kayıtları gibi pek çok veriyi işlediğinden veri madenciliği kullanır.

Perakende veri madenciliğinin kazanımları:

Müşteri alış davranışı belirlenmesi

Müşteri alışveriş örneklenmesi ve eylemin belirlenmesi

Müşteri servis kalitesinin artırılması

Müşteri hatırlama ve memnuniyetinin sağlanması

Ürün tüketim oranının artırılması

Ürün nakil ve dağıtım düzeninin tasarlanması

İş maliyetinin düşürülmesi

olarak sıralanabilir.

Ve Telekomünikasyon....



# Bilimsel ve İstatistiksel Veri Madenciliği

Kapsamlı bilimsel veriler için özellikle sayısal veriler üzerinde veri analizi için başvurulanan (örn:psikoloji, tıp, elektrik mühendisliği, ekonomi, sosyal bilimler) teknikler vardır. Ana istatistiksel metotlar şunlardır:

- Regresyon
- Genelleştirilmiş doğrusal modeller
- Regrasyon ağaçları (Regression trees)
- Uyuşmazlık analizi (Analysis of variance)
- Karma-etki modeli (Mixed effect model)
- Faktör analizi (Faktor analysis)
- Ayırım analizi (Disciriminant analysis)
- Zaman serileri
- Yaşam analizi (Survival analysis)
- Kalite kontrolü



# Kaynaklar

## Kaynaklar

- [1] Irma Becerra-Fernandez, Stelios H. Zanakis, Steven Walczak, “ Knowledge Discovery Techniques for Predicting Country Investment Risk,” **Computers& Industrial Engineering**, No: 43, 2002, s.787-800
- [2] Pete Chapman, v.d., “CRISP-DM 1.0 Step-by-Step Data Mining Guide,” (Çevrimiçi) <http://www.crisp-dm.org/CRISPWP-0800.pdf> , 12 Aralık 2004, s. 13.
- [3] Becker, Ghedini, **a.g.e.**, s. 97.
- [4] P.,Yıldırım,M.,Uludağ,,A.Görür,”Hastane Bilgi Sistemlerinde Veri Madenciliği”, Akademik Bilişim 2008, Çanakkale 18 Mart Üniversitesi.





















