

# VERİ MADENCİLİĞİ



# VERİ MADENCİLİĞİ



Veri Madenciliği, örüntü tanıma, istatistik ve matematiksel yöntemlerin kullanımıyla devasa miktardaki güncel ya da geçmiş veri içerisinde ilgi çekici ( **önemsiz olmayan, gizli, önceden bilinmeyen, potansiyel olarak kullanışlı** ) bilginin gelecekteki eğilimleri kestirmek ya da sonraki aşamalarda analiz etmek üzere etkin şekilde çıkarılması sürecidir.

# VERİ MADENCİLİĞİNİN AMACI



NE YAPMAK İSTEMİYORUZ ?

- ❧ Büyük miktardaki veri içinde arama yapmak (Veri tabanı yönetim sistemleri bu işi yapıyor)
  - ❧ Telefon rehberinde arama yapmak
- VERİ MADENCİLİĞİNİN AMACI
- ❧ Aradığımız veri mevcutsa bu veriden çıkarabileceğimiz sonuçlarını anlamak
  - ❧ Telefon rehberindeki isimlere göre , hangi bölgemizde hangi isimlerin daha yaygın olduğunu bulmak

# VERİ MADENCİLİĞİ

## VERİ TABANI

## VERİ MADENCİLİĞİ

### SORGULAMA



### SORGULAMA

- Tanımlı
- SQL

- Tam tanımlı değil
- Yaygın sorgulama dili yok

### VERİ

- Canlı veri
- Dinamik

### VERİ

- Üzerinde işlem yapılmayan veri

### ÇIKTI

- Belirli
- Verinin bir alt kümesi

- Statik

### ÇIKTI

- Belirli değil
- Verinin bir alt kümesi değil

# NEDEN VERİ MADENCİLİĞİ

---

- ❧ Bilgisayarların ucuzlayıp aynı zamanda çok güçlü hale gelmeleri
- ❧ Teknolojinin gelişimi ile bilgisayar ortamında ve veri tabanlarında tutulan veri miktarının da artması
- ❧ Yeni veri toplama yolları
  - Otomatik veri toplama aletleri, veri tabanı sistemleri, bilgisayar kullanımının artması
- ❧ Büyük veri kaynakları
  - İş dünyası : Web, e-ticaret, alışveriş , hisse senetleri
  - Bilim dünyası : uzaktan algılama ve izleme, simülasyonlar
  - Toplum : Haberler, dijital kameralar, youtube, facebook

# VERİ MADENCİLİĞİ



- ❧ Veri madenciliği tanımını ile birlikte bilinmesi gereken kavramlar :
- ❧ **Veri yığınları** : Soyut ve ilk bakışta anlaşılmayan veri türüdür.
- ❧ **Veri Ambarı** : Sınırı belirli olmayan ve ilişkisiz birçok verinin bulunduğu veri tabanlarıdır.

# VERİ MADENCİLİĞİ



## ∞ Anlamsız veriler arasında BİLGİ keşfi için

1. **Veri temizleme** (gürültülü ve tutarsız verileri çıkarmak)
2. **Veri bütünleştirme** (birçok veri kaynağını birleştirebilmek)
3. **Veri seçme** (yapılacak olan analizle ilgili olan verileri belirlemek )
4. **Veri dönüşümü** (verinin veri madenciliği tekniğinden kullanılabilir hale dönüşümünü gerçekleştirmek)
5. **Veri madenciliği** (veri örüntülerini yakalayabilmek için akıllı metotları uygulamak)
6. **Örüntü değerlendirme** (bâzı ölçümlere göre elde edilmiş bilgiyi temsil eden ilginç örüntüleri tanımlamak)
7. **Bilgi sunumu** (mâdenciliği yapılmış olan elde edilmiş bilginin kullanıcıya sunumunu gerçekleştirmek).

# VERİ MADENCİLİĞİ

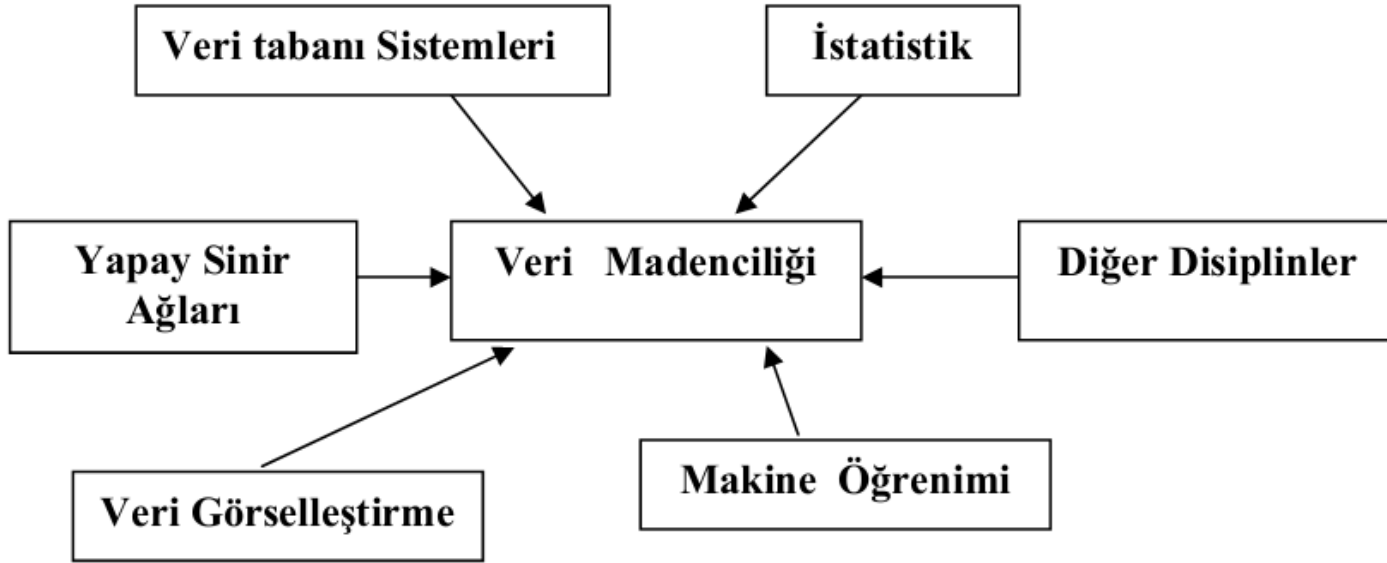


## Veri Madenciliği Sistemi

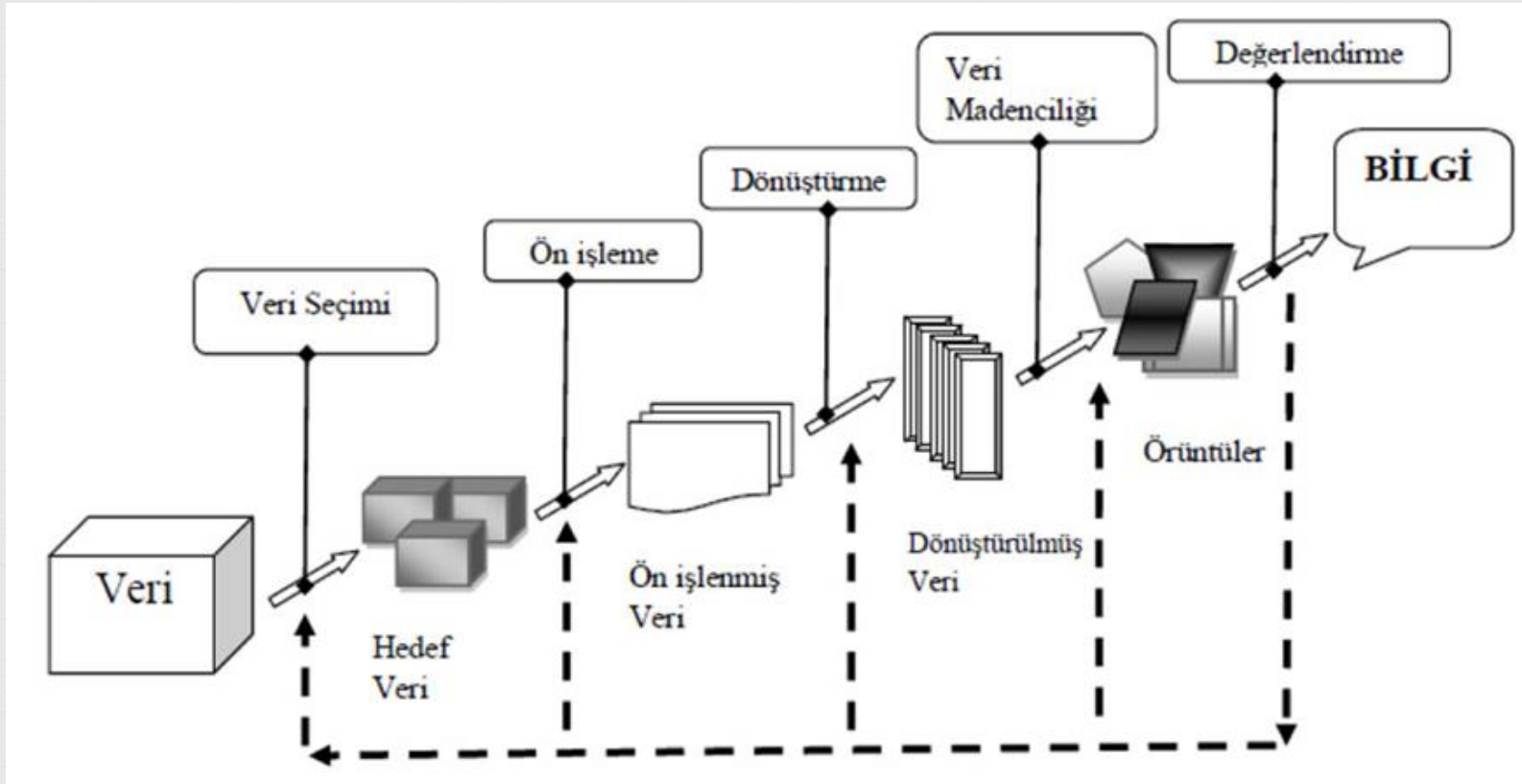
- Veritabanı, veri ambarı ve diğer depolama teknikleri
- Veritabanı ya da Veri Ambarı Sunucusu
- Bilgi Tabanı
- Veri Madenciliği Motoru
- Örüntü Değerlendirme
- Kullanıcı Arayüzü



# VERİ MADENCİLİĞİ



# VERİ MADENCİLİĞİ



# VERİ MADENCİLİĞİ



## 1 Veri Seçimi

- Veri seçimi, veri madenciliği aşamalarında en fazla zaman alan kısımlardan biridir. Bu aşamada bilgi sistemlerinde oluşan bilgi iyi analiz edilmelidir ve problemle ilişkilendirilmelidir. Analizi yapan kişinin veri kalitesini ölçmesi açısından bu aşama önemlidir. Büyük miktardaki verilerin tek bir veri tabanı veya veri ambarında birleştirilmesi veri madenciliği uygulaması için gereklidir. Veri seçimi aşaması filtreleme olarak da isimlendirilebilir.

## 2 Ön İşleme

- Ön işleme aşaması veri madenciliğinin başarısı için önemlidir. Bu aşamada veri, sonraki aşamalarda kullanılabilmesi için elverişli hale getirilir. Ön işleme aşamasının başarısı sonuçtaki başarıyı doğrudan etkiler. Başarılı bir ön işleme aşamasıyla kesin ve net sonuçlara ulaşmak mümkündür.

# VERİ MADENCİLİĞİ



## ❧ 3 İndirgeme

- ❧ Veri üzerinden faydalı ve doğru sonuç elde etmek için kullanılacak verinin indirgenmesi gerekir. Eldeki verinin büyük bir kısmı, her ne kadar ön işleme aşamasından geçmiş olsa bile sonraki aşamalarda kullanılacak durumda değildir. Dolayısıyla kullanılacak duruma indirgenmesi gerekir.

## ❧ 4 Veri Madenciliği

- ❧ Veri madenciliği çalışmasının tam olarak kullanıldığı aşamadır. Veri bu aşamaya gelince doğru ve kullanılabilir haldedir. Çalışmanın amacına göre bu aşamada veri madenciliği yöntemlerinden biri veya birkaçı uygulanır. Gerektiği durumlarda farklı yöntemler birleştirilerek kullanılabilir.

# VERİ MADENCİLİĞİ



## 5 Yorumlama ve Doğrulama

- Veri üzerinde veri madenciliği uygulandıktan sonra alınan sonuçlar yorumlanır ve çalışmanın doğru sonuca ulaşp ulaşmadığı araştırılır. Bu aşamada genellikle farklı yöntemler uygulanmışsa onların karşılaştırması yapılır. Elde edilen sonuçlar yapılmış olan diğer çalışmaların sonuçlarıyla karşılaştırılıp doğrulanır.

# VERİ MADENCİLİĞİ



Uygulama alanları

## ∞Pazarlama

- Müşterilerin satın alma örüntülerinin belirlenmesi,
- Müşterilerin demografik özellikleri arasındaki bağlantıların bulunması,
- Posta kampanyalarında cevap verme oranının artırılması,
- Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması,
- Pazar sepeti analizi (Market Basket Analysis)
- Müşteri ilişkileri yönetimi (Customer Relationship Management)
- Müşteri değerlendirme (Customer Value Analysis)
- Satış tahmini (Sales Forecasting).

# VERİ MADENCİLİĞİ



## Uygulama alanları

### Bankacılık

- Farklı finansal göstergeler arasında gizli korelasyonların bulunması,
- - Kredi kartı dolandırıcılıklarının tespiti,
- - Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi,
- Kredi taleplerinin değerlendirilmesi.

### Sigortacılık

- Yeni poliçe talep edecek müşterilerin tahmin edilmesi,
- Sigorta dolandırıcılıklarının tespiti,
- Riskli müşteri örüntülerinin belirlenmesi.

# VERİ MADENCİLİĞİ



∞ Veri Madenciliği süreci için sıkça kullanılan yazılımlar şunlardır:

- ❖ WEKA
- ❖ RapidMiner (YALE)
- ❖ R
- ❖ Orange,
- ❖ KNIME
- ❖ Tanagra



# VERİ MADENCİLİĞİ



## ☞ Neden WEKA

- ☞ makine öğrenmesi algoritmaları vardır
- ☞ veri setine doğrudan uygulanabilir
- ☞ kendi Java kodu içerisinde çağırılabilir
- ☞ veri işleme, sınıflandırma, regresyon, kümeleme, ilişki kuralları ve görüntüleme araçları içerir
- ☞ yeni makine öğrenmesi şemaları geliştirmek için uygun yapıdadır

# KAYNAKLAR



- ❧ Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2005). Introduction to Data Mining. Addison Wesley, ISBN:
- ❧ VERİ MADENCİLİĞİ DR. TUĞRUL TAŞCI
- ❧ David J. Hand, Heikki Mannila, and Padhraic Smyth (2001). Principles of Data Mining. MIT Press. ISBN X.

# SONUÇ



Veri madenciliđi, makina öğrenmesi ve yapay zeka birbiri yerine kullanılmakla birlikte, aslında farklı amaçları olan fakat ortak araçları kullanabilen bilimsel disiplinlerdir.



BEYZA YILDIRIM

201420404038