

VERİ MADENCİLİĞİ



FATMA İLHAN



Veri Madenciliği Nedir?



Veri madenciliği büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programı kullanılarak aranmasıdır.

Basit bir tanım yapmak gerekirse bilgiyi maden-leme işidir.





Veri madenciliđi; önceden bilinmeyen, geçerli ve uygulanabilir bilginin veri yığınlardan dinamik bir süreç ile elde edilmesi olarak tanımlanabilir. Bu süreçte; kümeleme, veri özetleme sınıflama kurallarının öğrenilmesi, bağımlılık ağlarının bulunması, deđişkenlik analizi ve anomali tespiti gibi farklı birçok teknik kullanılmaktadır.



Veri madenciliđi, teknolojinin hızla ilerlemesi ile büyük veri yığınlarından oluşan database sistemleri içerisinde gizli kalmıř bilgilerin çekilmesi sađlanır. Bu iřlem, istatistik, matematik disiplinleri, modelleme teknikleri, database teknolojisi ve çeřitli bilgisayar programları kullanılarak yapılır.

Veri Madenciliği Tarihi

Gelişim Adımları	Cevaplanan Karar Problemi	Kullanılabilen Teknolojiler	Ürün Sağlayıcıları	Karakteristikler
Veri Toplama (1960'lar)	"Benim toplam karımı geçen 5 yılda ne kadardı?"	Bilgisayarlar, Teypler, Diskler	IBM,CDC	Geriye dönük , statik veri dağıtımı
Veri Erişimi (1980'ler)	"İngiltere'de geçen mart ayında birim satışları ne kadardı?"	İlişkisel Veritabanları, SQL, ODBC	Oracle,Sybase, Informax,IBM, Microsoft	Kayıt düzeyinde geriye dönük, dinamik veri dağıtımı
Veri Ambarlama ve Karar Destek Sistemleri (1990'lar)	"İngiltere'de geçen mart ayında birim satışları ne kadardı?"	OLAP, Çok Boyutlu Veritabanı Sistemleri, Veri ambarları	Pilot, Comshare, Arbor,Cognos, Microstrategy	Çoklu düzeylerde, geriye dönük dinamik veri dağıtımı
Veri Madenciliği (Bugün)	"Gelecek ay Boston'daki birim satışlar muhtemelen ne olabilir, niçin?"	İleri düzeyde algoritmalar, çok işlemcili bilgisayarlar, büyük veritabanları	Pilot, Lockheed, IBM,SGL, SPSS,SAS, Microsoft vs.	Geleceğe dönük ,proaktif enformasyon dağıtımı

Neden Bu Kadar Önemli?



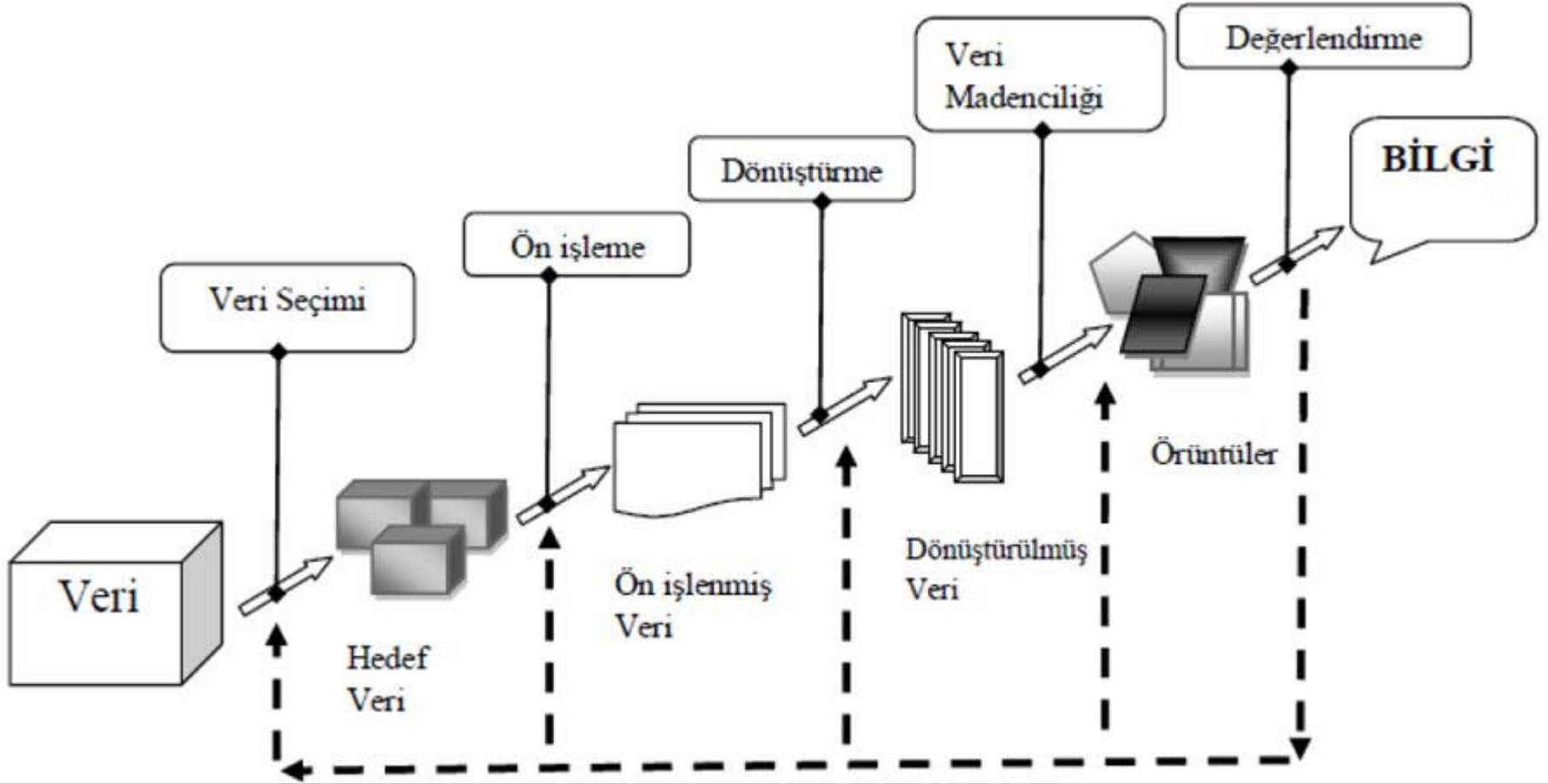
1995 yılında birincisi düzenlenen Veri Tabanlarında Bilgi Keşfi Konferansı bildiri kitabı aşağıdaki ifadelerle başlamaktadır :

“Dünyadaki enformasyon miktarının her 20 ayda bir ikiye katlandığı tahmin edilmektedir. Bu ham veri seli ile ne yapmamız gerekmektedir. İnsan gözleri bunun ancak çok küçük bir kısmını görebilecektir.”



Veri madenciliđi son yıllarda ok yođun bir Őekilde bilgi endüstrisinin ilgisini ekmektedir. Bu ilginin temel nedeni ise ok büyük miktarda verilerin elde edilebilmesi ve elde edilen bu ham verilerin hızlı bir Őekilde faydalı bilgilere dönüŐtürölmesi gerekliliđidir.

Veri Madenciliği Aşamaları



1. Veri Seçimi(Data Selection)

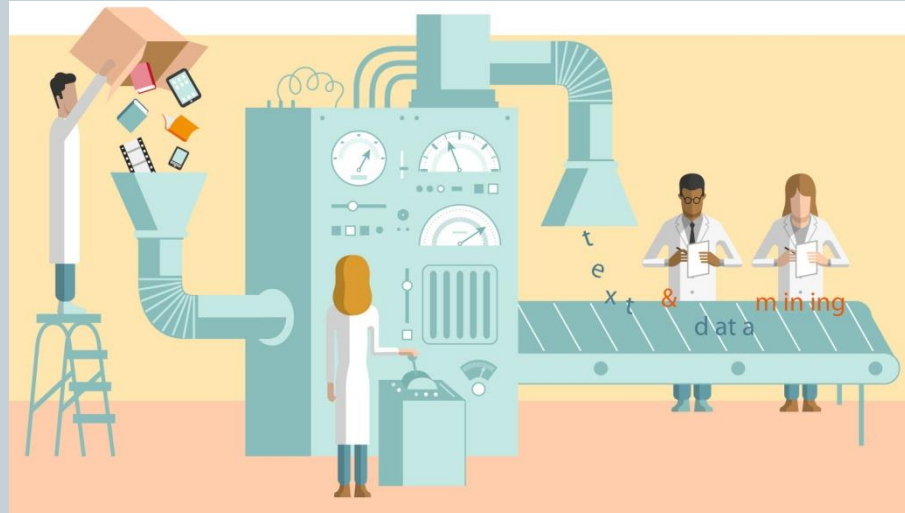


Veri seçimi, veri madenciliği aşamalarında en fazla zaman alan kısımlardan biridir. Bu aşamada bilgi sistemlerinde oluşan bilgi iyi analiz edilmelidir ve problemle ilişkilendirilmelidir. Büyük miktardaki verilerin tek bir veri tabanı veya veri ambarında birleştirilmesi veri madenciliği uygulaması için gereklidir.

2. Ön İşleme(Preprocessing)



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



3. İndirgeme(Data Reduction)

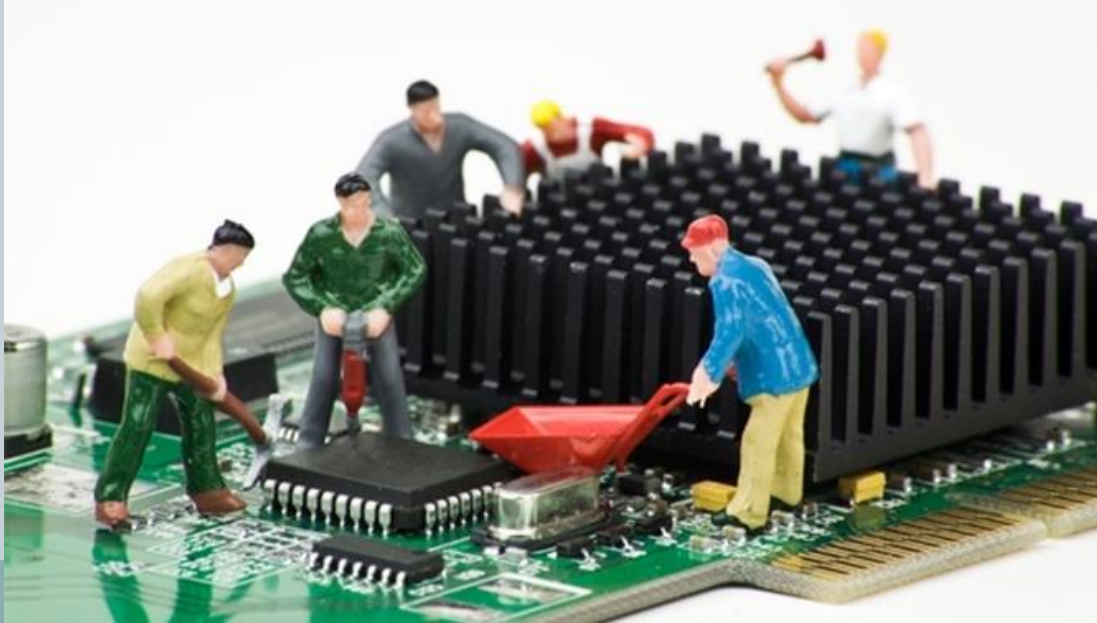


Veri üzerinden faydalı ve doğru sonuç elde etmek için kullanılacak verinin indirgenmesi gerekir. Eldeki verinin büyük bir kısmı, her ne kadar ön işleme aşamasından geçmiş olsa bile sonraki aşamalarda kullanılabilir durumda değildir.

4. Veri Madenciliđi(Data Mining)



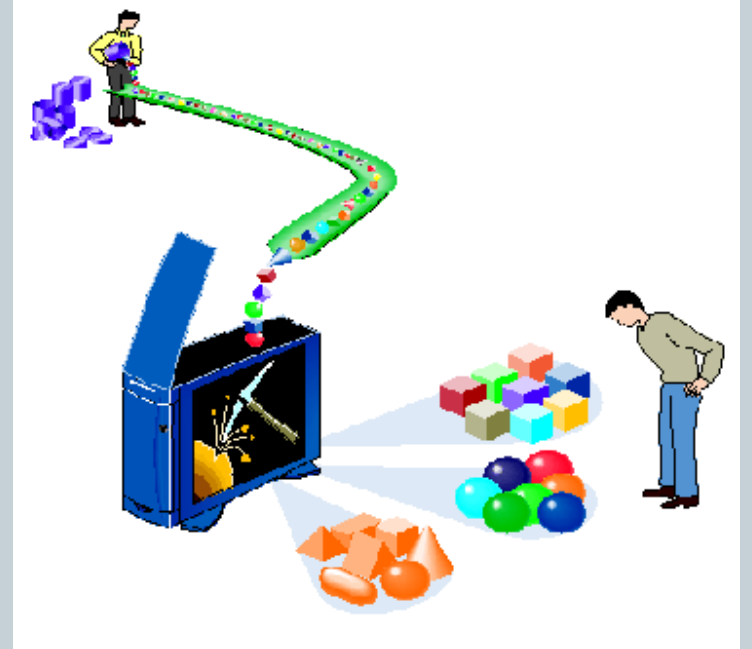
Veri madenciliđi alıřmasının tam olarak kullanıldıđı ařamadır. Veri bu ařamaya gelince dođru ve kullanılabilir haldedir. alıřmanın amacına gre bu ařamada veri madenciliđi yntemlerinden biri veya birkaçı uygulanır.



5. Deęerlendirme (Evaluation)



Veri üzerinde veri madencilięi uygulandıktan sonra alınan sonuçlar yorumlanır ve alıřmanın doęru sonuca ulařıp ulařmadıęı arařtırılır. Bu ařamada genellikle farklı yntemler uygulanmıřsa onların karřılařtırılması yapılır. Elde edilen sonuçlar yapılmıř olan dięer alıřmaların sonuçlarıyla karřılařtırılıp doęrulanır.



Veri Madenciliği Yöntemleri

Veri Madenciliği Yöntemleri

Öngörü Yöntemleri

Sınıflandırma (Classification)

Karar Ağaçları
(Decision Trees)

Bayes Sınıflandırması
(Bayesian Classification)

En Yakın Komşu
(Nearest Neighbour)

Yapay Sinir Ağları
(Neural Networks)

Karar Destek Makineleri
(Support Vector Machines)

Zaman Serisi Analizi
(Time Series Analysis)

Diğer Yöntemler

Eğri Uydurma (Regression)

Tanımlayıcı Yöntemler

Demetleme (Clustering)

Birliktelik Analizi
(Association Analysis)

Sıralı Dizi Analizi
(Sequence Analysis)

Özetleme
(Summerization)

Tanımsal İstatistik
(Descriptive Statistics)

İstisna Analizi
(Outliner Analysis)

Diğer Yöntemler

Öngörü Yöntemleri



Öngörü Yöntemleri sonuçları bilinen verilerden hareket ederek bir model geliştirilmesi ve bu modelden faydalanılarak sonuçları bilinmeyen veriler için tahmin etme amaçlanır.

- Örneğin ; Bir bankanın , ilk 3 taksitinden 2 veya daha fazlasını geç ödeyen müşterileri için %75 'inin krediyi geri ödeyememesine dayalı tahmini bu modele örnek olarak verilebilir.

Tanımlayıcı Yöntemler



Tanımlayıcı yöntemler ise mevcut verilerin tanımlanmasını sağlamak olarak ifade edilebilir.

- Geliri A-B aralığında ve arabası olan çocuklu aileler ile geliri A-B aralığından düşük çocuđu olmayan ailelerin satın alma örüntülerinin birbirine benzerlik göstermesinin saptanması bu modele örnek olarak verilebilir.



Şimdi ise veri madenciliğinin bu yöntemlerini işlevlerine göre sınıflandıralım.

- Sınıflama ve Regresyon
 - Kümeleme
 - Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler
- olarak 3 ana başlıkta incelemek mümkündür. Sınıflama ve Regresyon yöntemleri öngörücü; Kümeleme, Birliktelik Kuralları ve Ardışık Zamanlı Örüntü yöntemleri ise tanımlayıcı yöntemlerdir.

Sınıflama:



Veriyi önceden belirlenmiş sınıflardan birine dahil etmedir.

- *Örneğin :*

Kızlara pembe, erkeklere mavi kıyafet gibi...

Regresyon :



Veriyi gerçel değerli bir fonksiyona dönüştürmedir.

- *Örneğin :*

Vizeden 90 alan dersten geçecek veya evi ve arabası olan krediyi öder gibi...

Kümeleme :



Benzer verileri aynı grupta toplamadır.



Birliktelik Kuralları :



Eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır.

- *Örneğin :*

Müşteriler sigara satın aldığıında %50 ihtimalle çakmakta satın alırlar .

Ardışık Zamanlı Örüntüler :



Birbirleri ile ilişkisi olan ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılır.

- *Örneğin :*

X ameliyatı yapıldığında 15 gün içinde %45 ihtimalle Y enfeksiyonu oluşacaktır.

Veri Madenciliđi Uygulama Alanları



- Finans: Kredi analizlerinde
- Sigorta: Sahtekarlık Analizi, Hasar tespiti
- Telekomünikasyon: Arama kayıtlarının analizi
- Müşteri tüketimleri: Promosyon analizlerinde
- Bilimsel araştırmalar: Resim, video, konuşma
- Kamu hizmetlerinde: Enerji kullanım analizlerinde
- Sağlık ve Biyoloji: DNA verilerinin analizi
- Taşımacılık: Lojistik yönetimi

KAYNAKÇA



- <https://industryolog.com/veri-madenciligi-data-mining/> - Kardelen Bal
- VERİ MADENCİLİĞİ; Veriden Bilgiye, Masraftan Değere - Dr. Yılmaz ARGUDEN ve Burak ERŞAHİN
- Veri Madenciligi: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları, Ali Serhan Koyuncugilı ve Nermin Özgülbaş, BİLİŞİM TEKNOLOJİLERİ DERGİSİ, CİLT: 2, SAYI: 2, MAYIS 2009
- <http://mgocenoglu.blogspot.com.tr/2014/06/veri-madenciligi-asamalar.html> - Veri Madenciligi Aşamaları (son erişim: 26.11.2017)

TEŐEKKÜRLER



FATMA İLHAN
BALIKESİR ÜNİVERSİTESİ
ENDÜSTRİ MÜHENDİSLİĐİ
201410404074