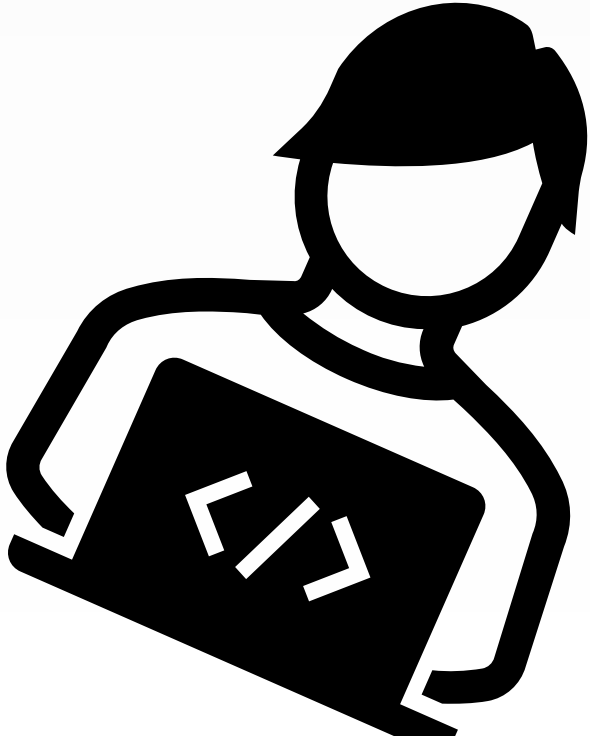
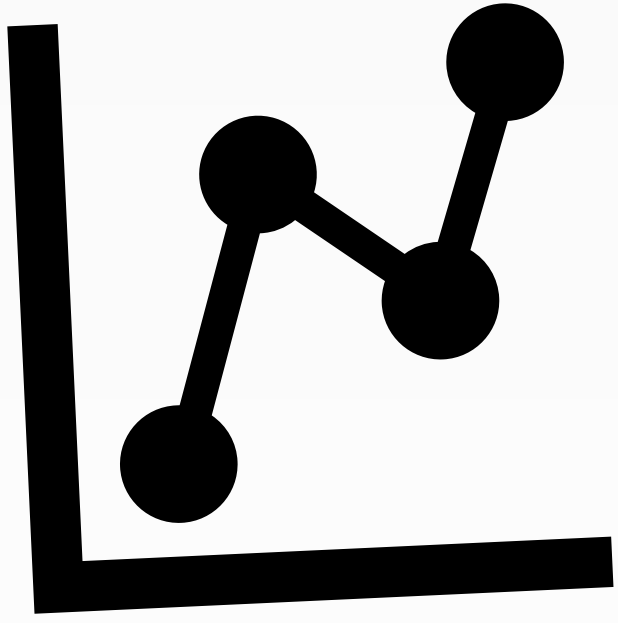




EKO 469-Veri Madenciliği

Hafta 1 - Giriş

Ders Hakkında

DERS HAKKINDA

- Hızla teknolojinin hayatımıza girmesiyle elektronik ortamlarda tutulan verilerin artması, verilerin nasıl ve nerelerde kullanılacağı, nasıl yorumlanacağı ve bilgiye nasıl ulaşılacağı ihtiyacını ortaya çıkarmıştır.
- Gelişen teknoloji ile birlikte kullanıcı talepleri de artarak çeşitlenmiştir. Bu itibarla basit raporlamalar ve standart sorguların yeterli gelmemesi de sektör çalışanlarını arayışa itmiştir.
- **Veri yönetimi, kuruluşların en temel faaliyetlerinden biridir. Doğru, ilgili ve zamanında bilgi, karar verme süreçlerinde çok etkilidir ve kuruluşların hayatta kalması için son derece önemlidir.**

Bu dersin temel çıktısı, bilişim teknolojileri dünyasındaki önemini her geçen gün daha da arttıran veri madenciliği konusunu ve veri madenciliği modellerini öğrenmektir.

İçerik Örnekleri

RapidMiner ve Python yazılım dili kullanılarak veri madenciliği uygulamaları geliştirilecektir.



Ders Akışı



Veri Madenciliği Temelleri



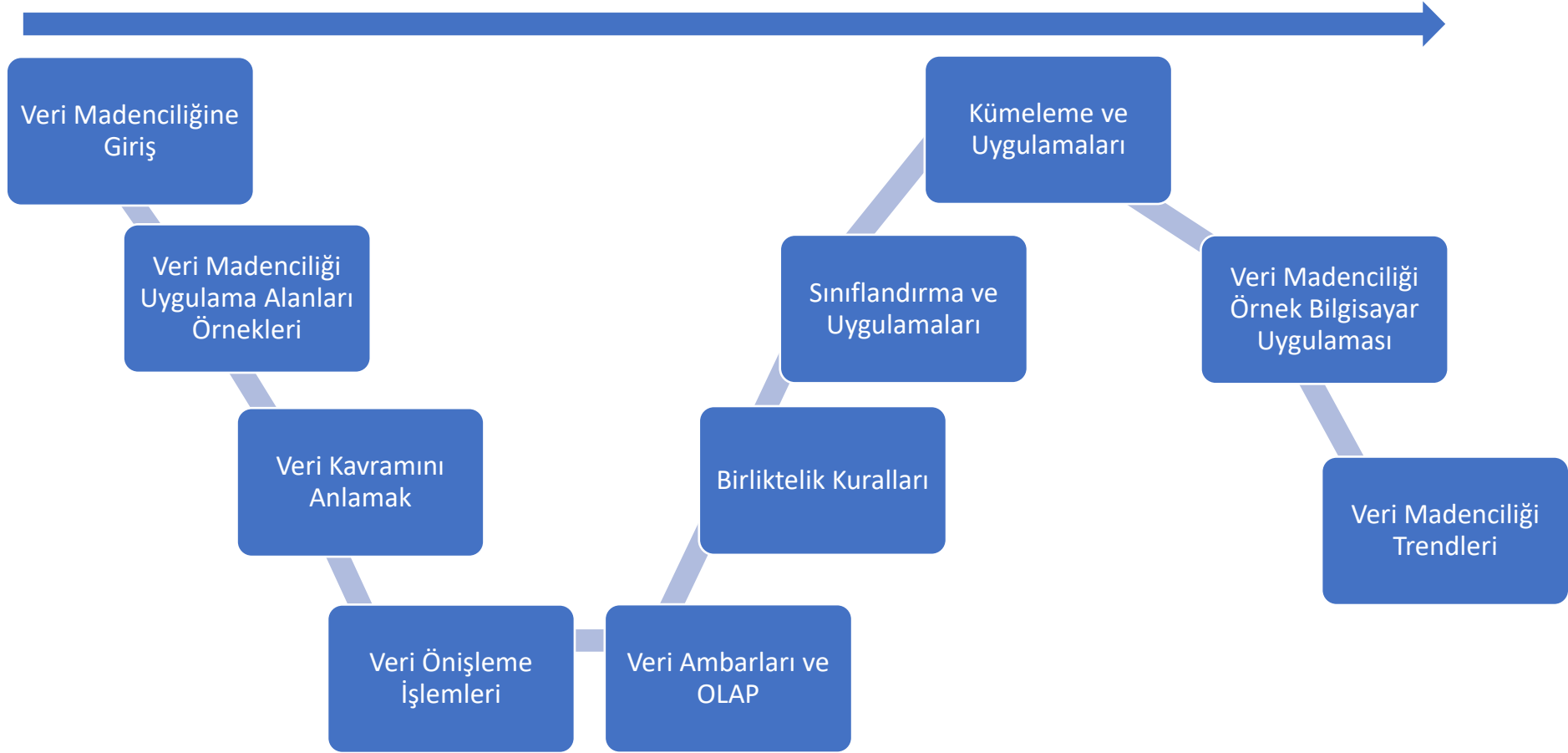
Önişleme



Birliktelik, Sınıflandırma
ve Kümeleme



Uygulamalar ve Trendler



Değerlendirme Sistemi

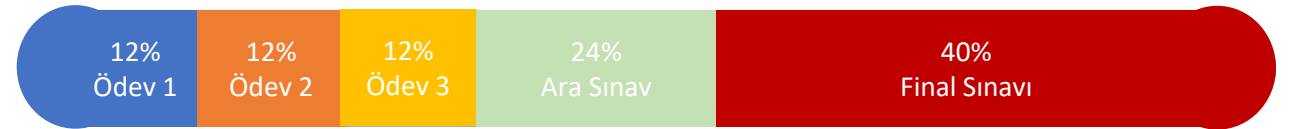
•Yıliçi Çalışmaları

- Ödev 1 - %20
- Ödev 2 - %20
- Ödev 3 - %20
- Ara Sınav (Vize) - %40

//Detayları paylaşılacak
//Detayları paylaşılacak
//Detayları paylaşılacak

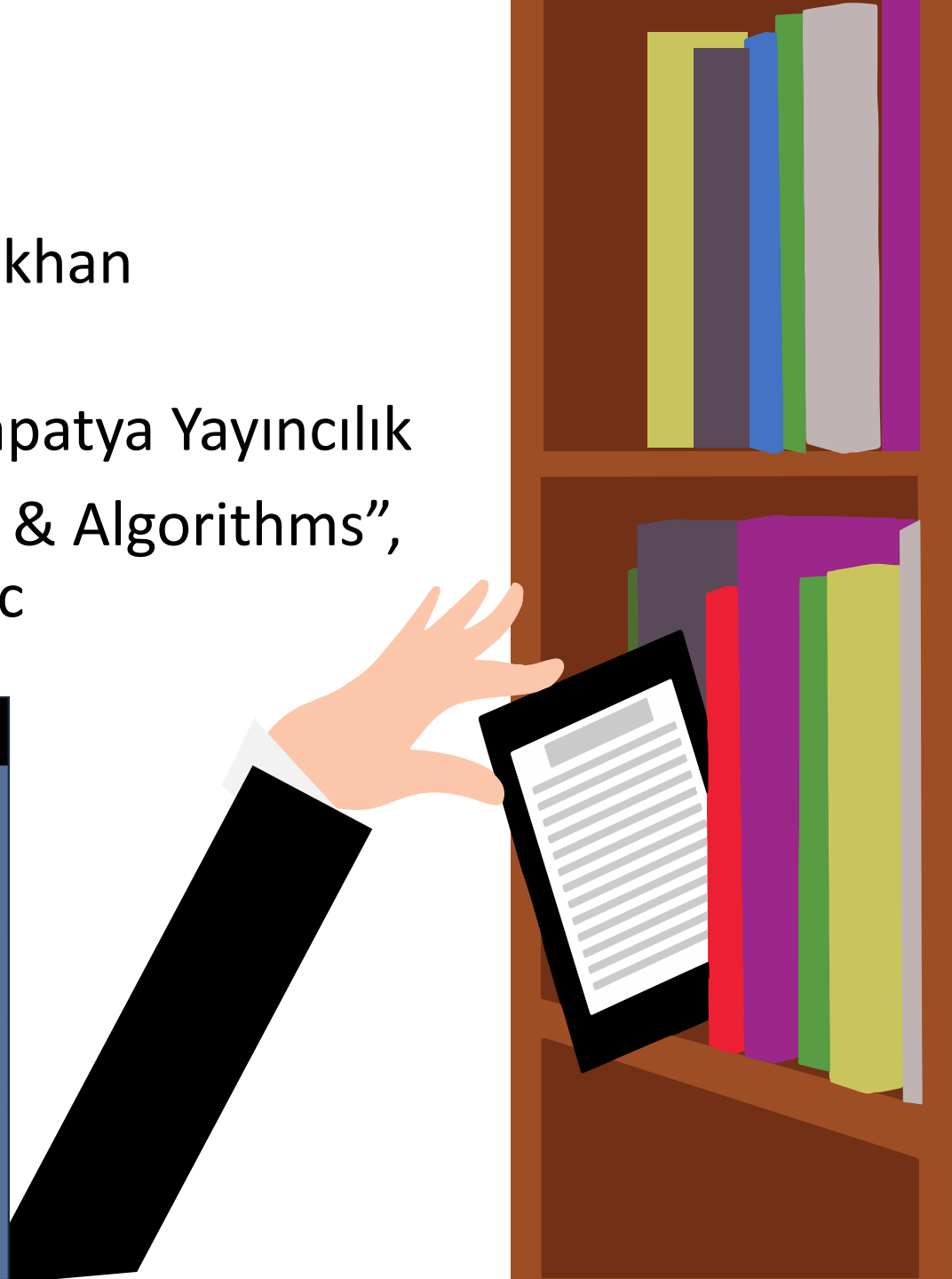
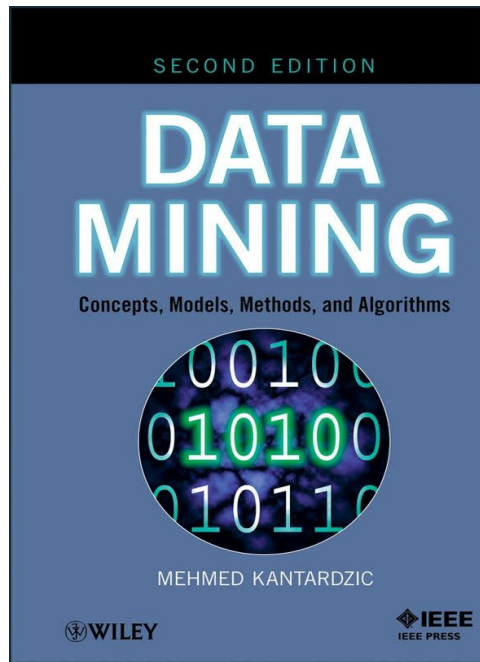
•Final Değerlendirmesi

- Yıliçi - %60
- Final Sınavı - %40



Temel Kaynaklar

- 1- Veri madenciliği Kavram ve Algoritmaları, Gökhan Silahtaroglu, Papatya Yayıncılık
- 2- Veri Madenciliği Yöntemleri, Yalçın Özkan, Papatya Yayıncılık
- 3- “DATA MINING: Concepts, Models, Methods & Algorithms”, Second Edition, John Wiley, 2011, M. Kantardzic



Tartışma Konuları

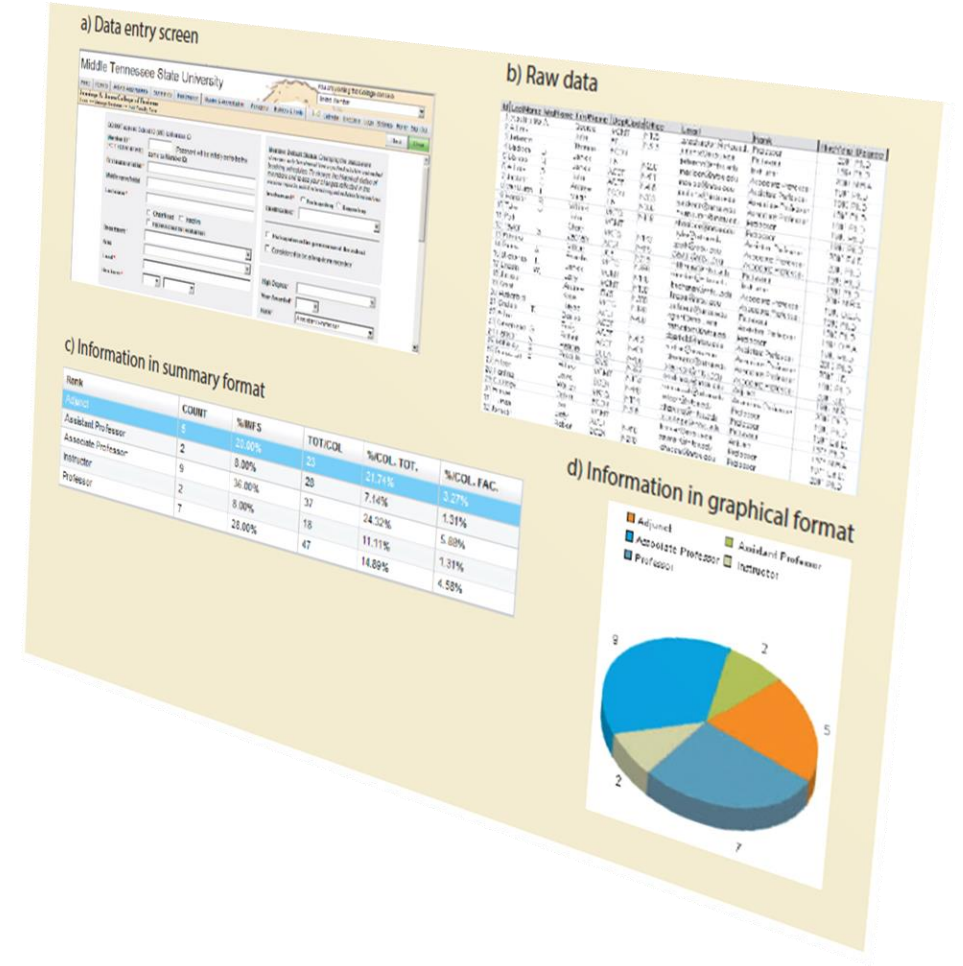
Tartışma Konuları

- Veri ve Bilgi kavramları arasındaki fark nedir?
- Veri Madenciliği nedir? Ne değildir?
- Veri Madenciliği kullanım alanları
- Veri Madenciliği önemi
- Veri Madenciliği süreçleri



Veri, Bilişim, Bilgi ve Bilgelik

- **Veri (Data):** Tek başlarına bir anlam ifade etmeyen ham gerçekliktir. (Semboller ve sinyaller)
- **Bilişim (Information):** Verinin biçim kazandırılmış /anlamlandırılmış halidir.
- **Bilgi (Knowledge):** Bilişimin kişisel/kurumsal kararlara uygulanabilir hale dönüştürülmüş halidir.
- Bilişimin bilgiye dönüşmesi, bireyin onu *algılaması, özümsemesi ve sonuç çıkarmasıyla* gerçekleşir.
- **Bilgelik (Wisdom):** Bilginin nedeninin ve ilişkisel bağlantılarının öğrenilmesidir. Yetenek, tecrübe gibi kişisel nitelikler birer bilgelik elemanıdır.
- Veri, bilişim ve bilgi gerçekleşmiş olgu ile, bilgelik gerçekleşebilecek olgularla alakalıdır.



Veri, Bilişim, Bilgi ve Bilgelik

Veri (Data)

- Ham gerçekler; enformasyonun inşa blokları
- İşlenmemiş bilgi

Enformasyon-Bilişim (Information)

- Anlamı ortaya çıkaracak şekilde işlenen veriler.

Doğru, ilgili ve zamanında enformasyon iyi bir karar vermenin anahtarıdır.
İyi bir karar verme ise işletmeler için global çevrede hayatta kalmanın anahtarıdır.

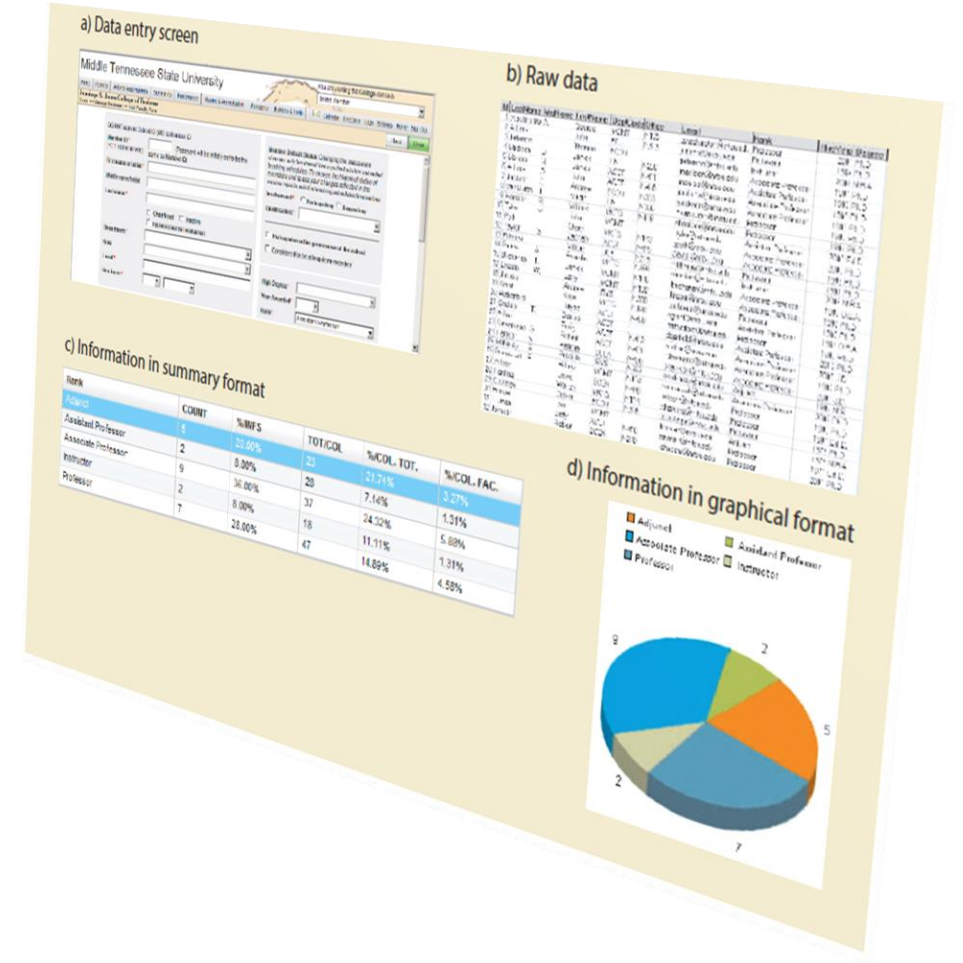
19 – genel bağlamını belirtmediğimiz sürece sizin için pek bir mana ifade etmez (VERİ)
Yandaki sayı yıllık enflasyon yüzdesini göstermektedir. – (ENFORMASYON)

Enformasyon, karar vermenin temeli olarak kullanılabilir. Örneğin, anket formundaki her soru için veri özeti, müşterilerin ihtiyaçlarını daha iyi karşılamak için bilinçli kararlar almanıza yardımcı olarak güçlü ve zayıf yönleri gösterebilir.

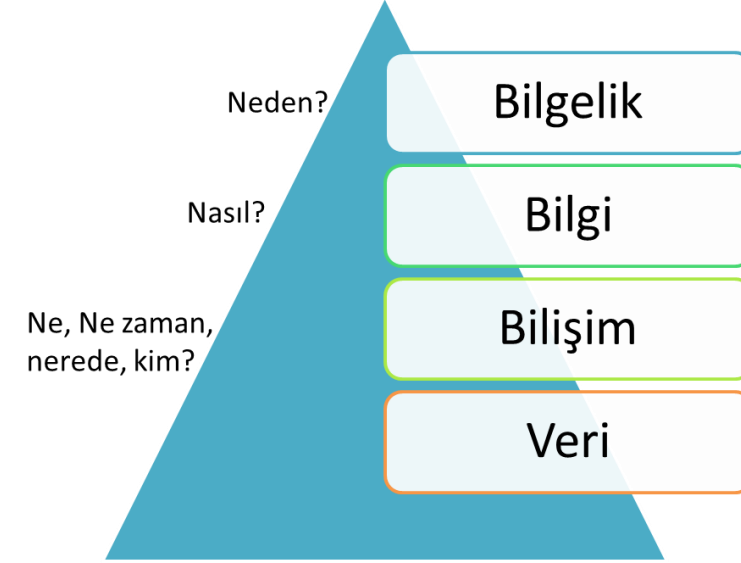
Bilgi (Knowledge)

- Bilgi ise enformasyona dönüştürülmüş verilerin biçimlendirilmiş ve anlamlandırılmış halidir. Bilgi, bir ortama uyarlanmış olan enformasyonun aşinalık, farkındalık ve anlayışını ifade eder. Bilginin anahtar bir karakteristiği yeni bilginin eski bilgi ve tecrübelerden türetilmesidir.

Verilerden elde edilen enformasyon/bilgilere dayanarak kariyeriniz boyunca kararlar alacaksınız. Bu nedenle, bu kavramlar arasındaki farkı bilmeniz sizler için önemlidir.



Veri, Bilişim, Bilgi ve Bilgelik



Kavram	Açıklama	Örnek
Bilgelik	Bilgi + Öngörü + Hareket => Eyleme geçilebilir İstihbarat	•Dursam iyi olacak!!!
Bilgi	Bilişim + Anlam => İlişkiler	•Şu anda yaklaşımakta olduğum trafik ışığı kırmızı yanıyor
Bilişim	Veri + Kapsam => Açıklama	•Trafik lambası: Kırmızı •Yer: Serdivan Gel-tat Yanı
Veri	Bilişim Temeli => Gözlem	•Kırmızı •40,762407 – 30,363087

Peki ya Siz?

Günlük yařantınızdan veri kullanım örnekleri verebilir misiniz?



Peki ya Siz?

- İşletmeler bu kadar veriyi nasıl ve ne amaçla işleyecekler?



SABİS

Öğrenci bilgileriniz
Seçilen dersleriniz ve notlarınız
Ders ve öğretim üyeleri arasındaki ilişki



Süper Market

Ürün verisi
Alım sonrası stoktaki ürün miktarı
Ödemeyi ne ile yaptınız



Online Alış-Veriş

Ürün ve stok verileri
Ürün yorumları
Kredi Kartı bilginiz



Seyehat

Seyehat acentası uçak ve otel bilgileri
Web sayfasında hangi müşteri bilgileri alınacak ve nerede tutulacak



Banka

Müşteri bilgileriniz, hesap numaranız ve bakiye bilgileriniz

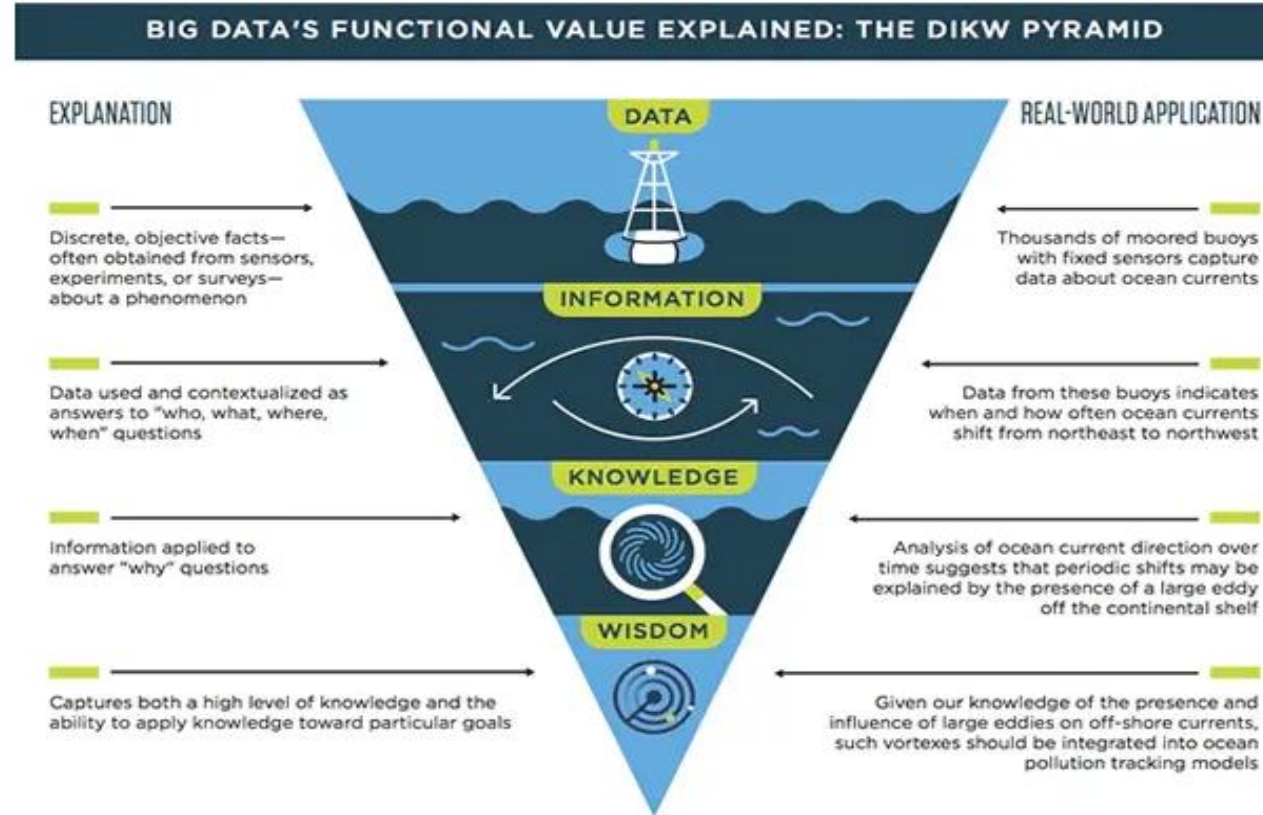


Sosyal Hesap

Arkadaşlarınız ve gruplarınız hakkındaki bilgiler
«Beğeniler(likes)»



DIKW Piramit



Büyük Veri üzerinde yukarıdaki açıklama ve örnekleri inceleyelim



Özet Olarak:

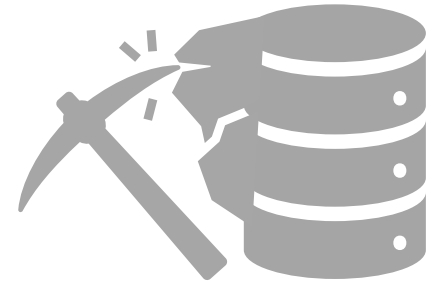
- Veriler, enformasyonun (bilşim) yapı taşlarını oluşturur.
- Enformasyon, verilerin işlenmesi sonucu üretilir.
- Enformasyon verinin anlamının ortaya çıkarılmasında kullanılır.
- Doğru, ilgili ve zamanında enformasyon iyi karar vermenin anahtarıdır.
- İyi karar verme global çevrede hayatta kalmanın anahtarıdır.
- Veriler uygun şekilde oluşturulmalı, erişilmesi ve işlenmesi kolay bir formatta saklanmalıdır.

Veri yönetimi; verilerin uygun şekilde üretilmesine, depolanmasına ve alınmasına odaklanan bir disiplindir. Herhangi bir işletme, devlet kurumu veya hizmet kuruluşu için temel niteliğinde bir aktivitedir.



Veri Madenciliği nedir?

- Veri madenciliği, büyük ölçekli veriler arasından değerli olan bir bilgiyi elde etme işidir.
- Verilerden üstü kapalı, çok net olmayan , önceden bilinmeyen ancak potansiyel olarak kullanılabilir geçerli ve uygulanabilir bilgilerin ve örüntülerin çıkarılması olarak tanımlanabilir.
- Biriken veri içerisinde yararlı olanlarını bulup ortaya çıkarma işine **veri madenciliği** adı verilir.
- Veri madenciliği daha önce hiç akla gelmemiş, düşünülmemiş sonuçları önümüze koymasıyla diğer yöntemlerden farklılık gösterir. Bu açıdan bakıldığında kurumların karar destek sistemleri için önemli bir yere sahiptir.



Neden Veri Madenciliği?

- Veri toplama ve veri depolama teknolojilerindeki hızlı gelişmeler, kuruluşların büyük miktarda veri biriktirmesine olanak sağlamıştır.
- Geleneksel veri analizi araçları ve teknikleri, veri kümesinin çok büyük boyutlarda olması nedeniyle veri analizi ve yararlı bilgilerin çıkarılması için kullanılamaz.
- İşletme, bilimsel ve İnternet verileri bugün yüzlerce terabayt olarak ölçülmektedir ve bu eğilim, görüntü, video ve ses gibi yeni multimedya veri türlerinin bu veri akışını daha da hızlı hale getirmesiyle katlanarak artmaktadır.

Günümüzde tüm veriler elektronik ortamda tutuluyor, peki bu veriler ne olacak? Biriken veriler bilgiye dönüştürülebilecek mi?

Günümüz verileri; Terabayt (1024 gigabayt), Petabayt, Exabayt, Zetabayt, Yotabayt boyutlarında.

- Google saniyede ortalama 40.000 arama işliyor (günlük ortalama 3,5 milyar, toplam günlük 5 milyar arama. Arama sonuçları kullanıcıya hızlı bir şekilde getirilebilir. (2018)
- Facebook kullanıcı sayısı 2 milyar. Her gün ortalama 1,5 milyar kullanıcı aktiftir. (2018)
- Her gün ortalama 40.000 yeni şarkı Spotify yükleniyor.
- Her dakika; 4.146.600 YouTube videosu izlendi, 456.000 tweet atıldı, Instagram'a 46.740 fotoğraf yüklendi, Facebook'a 510.000 yorum eklendi. (2018)
- Sprint ve AT&T gibi telekomünikasyon şirketlerinin, trilyonlarca telefon görüşmesinde verileri tutan sistemlere sahip olduğu biliniyor ve sisteme saniyede 70.000 çağrıya varan hızlarda yeni veriler ekleniyor.



Neden Veri Madenciliği?

- Büyük, karmaşık, bilgi açısından zengin veri kümelerini anlama ihtiyacı, neredeyse tüm işletme, bilim ve mühendislik alanlarında ortaktır.
- İş dünyasında, kurumsal ve müşteri verileri stratejik bir varlık olarak kabul edilmektedir. Bu verilerde saklı olan faydalı bilgileri çıkarma ve bu bilgiler üzerinden hareket etme yeteneği, günümüzün rekabetçi dünyasında giderek daha önemli hale gelmekte. Ancak sorun, bu büyük miktarda verinin toplanması ve düzenlenmesinde değildir. Sorun şu ki, bu petabaytlarca veya eksabaytlarca veriyi analiz etmek için uygun teknolojiye sahip değiliz.
- Sonuçlar:
 - Toplanan verilerin sadece küçük bir kısmı (%1 - %10) analiz edilmekte.
 - Hiç analiz edilemeyecek veriler büyük masraflarla toplanmaya devam ediliyor.

VERİ İÇERİSİNDE YÜZÜYÖRÜZ FAKAT BİLGİ AÇLIĞINDAN ÖLÜYÖRÜZ.



Neden Veri Madenciliği?

- İşletme, tıp, bilim veya devlet olsun, veri kümelerinin kendileri, ham formlarında çok az doğrudan değere sahiptir.
- Değerli olan, verilerden çıkarsanabilen ve kullanıma sunulabilen bilgidir.
- Günümüzde Veri madenciliğinin parlaması;
 - Analizler için büyük miktarda veri mevcut,
 - Veri toplama, düzenleme, arama ve dağıtma teknolojilerinin kullanılabilir hale gelmesi ve olgunlaşması,
 - İş dünyasında ve bilim camiasındaki rekabet baskısı
 - Başlangıçta sadece araştırma amaçlı geliştirilen bazı algoritmaların olgunlaşması ve ticari kullanıma açılması.



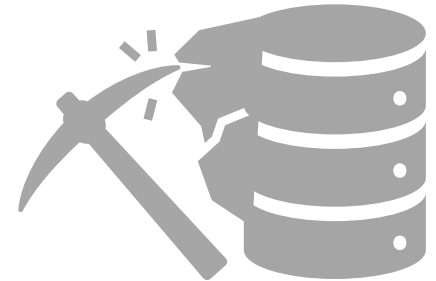
Veri Madenciliđi nedir?

- Mercedes sahibi kişiler ne tür özelliklere sahip?
- Bu müşteri için kredi geri ödemesi ne kadar olmalı?
- İşletme için iyi / yeni müşteri kimdir?
- Bu makaleye benzeyen başka makaleler var mı?
- Borsa indeksinin değeri önümüzdeki ay ne olabilir?
- Otomobil alan müşterinin hangi öznitelikleri önemlidir?
- En çok bir arada satılma olasılığı olan ürünlerimiz hangileridir?



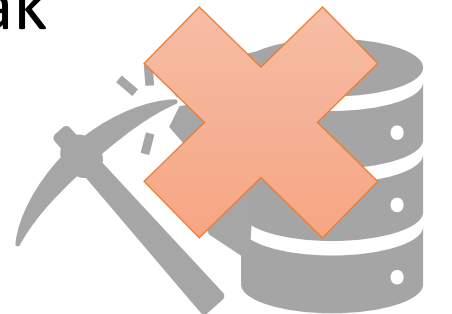
Veri Madenciliği nedir?

- Enerji tüketimimiz önümüzdeki ay ne kadar olabilir?
- Hangi ürün promosyonlarının kârlılık üzerindeki etkisi en yüksek?
- En iyi ürün dağıtım kanalı hangisi?
- Semptomlar ve hastalıklar arasındaki ilişkilerin keşfi
- Keşfedilen bu yeni canlı hangi sınıfa ait?
- Market raflarındaki ürünler nasıl dizilmeli?
- ATM'de günlük olarak ne kadar para tutulmalı?



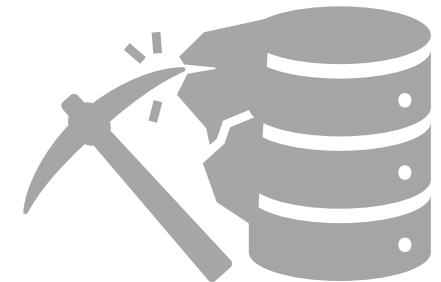
Veri Madenciliği ne değildir?

- Bir telefon defterinden telefon numarası aramak
- Arama motorlarından anahtar kelime aramak
- Maaşların farklı yaş gruplarına göre dağılım grafiğini çıkarmak
- Bir SQL sorgusuyla veritabanından sonuç döndürmek
- İlişkisel bir veritabanından çok boyutlu veri küpleri oluşturmak



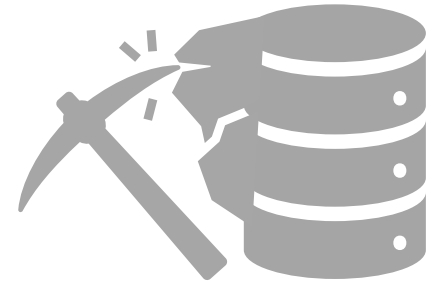
Veri Madenciliği Kullanım Alanları

- ☐ Pazarlama
- ☐ Üretim
- ☐ Bankacılık
- ☐ Finans
- ☐ Sigortacılık
- ☐ Elektronik Ticaret
- ☐ Eğitim-Öğretim
- ☐ Tıp
- ☐ Biyoloji
- ☐ Astronomi
- ☐ ...



Veri Madenciliği Kullanım Alanları

- ☐ İşletmelerde verimlilik / karlılık artışı
- ☐ Giderlerinin azaltılması
- ☐ Zarar oluşmadan tahmin edip ortadan kaldırılması
- ☐ Risk Yönetimi / Hilekarlık Tespiti
- ☐ Bilimsel araştırmalarda hız ve etkinlik artışı
- ☐ İnsansız sistemlerin gelişimine destek
- ☐ Eğitim – Sağlık – Güvenlik ve diğer birçok sektörde önleyici tedbirler alınması ve hizmet iyileştirme
- ☐ Kaynak / Enerji yönetimi
- ☐ ...



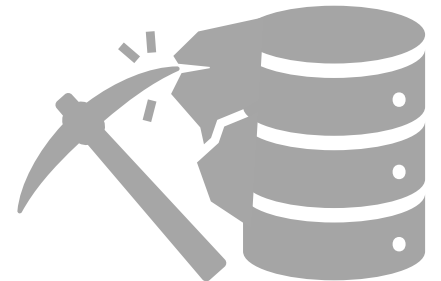
Örnek

Possible Business Discoveries

Acme Investors Incorporated

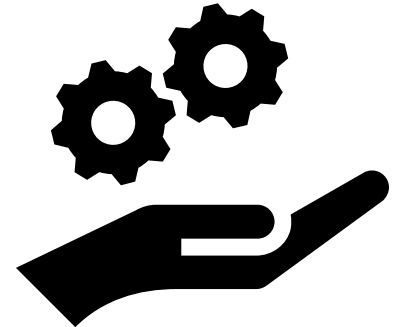
Customer ID	Account Type	Margin Account	Transaction Method	Trades/ Month	Sex	Age	Favorite Recreation	Annual Income
1005	Joint	No	Online	12.5	F	30-39	Tennis	40-59K
1013	Custodial	No	Broker	0.5	F	50-59	Skiing	80-99K
1245	Joint	No	Online	3.6	M	20-29	Golf	20-39K
2110	Individual	Yes	Broker	22.3	M	30-39	Fishing	40-59K
1001	Individual	Yes	Online	5.0	M	40-49	Golf	60-79K

- Can I develop a general characterisation/profile of different investor types? (**CLASSIFICATION**)
- What characteristics distinguish between Online and Broker investors? (**DISCRIMINATION**)
- Can I develop a model which will predict the average trades/month for a new investor? (**PREDICTION**)

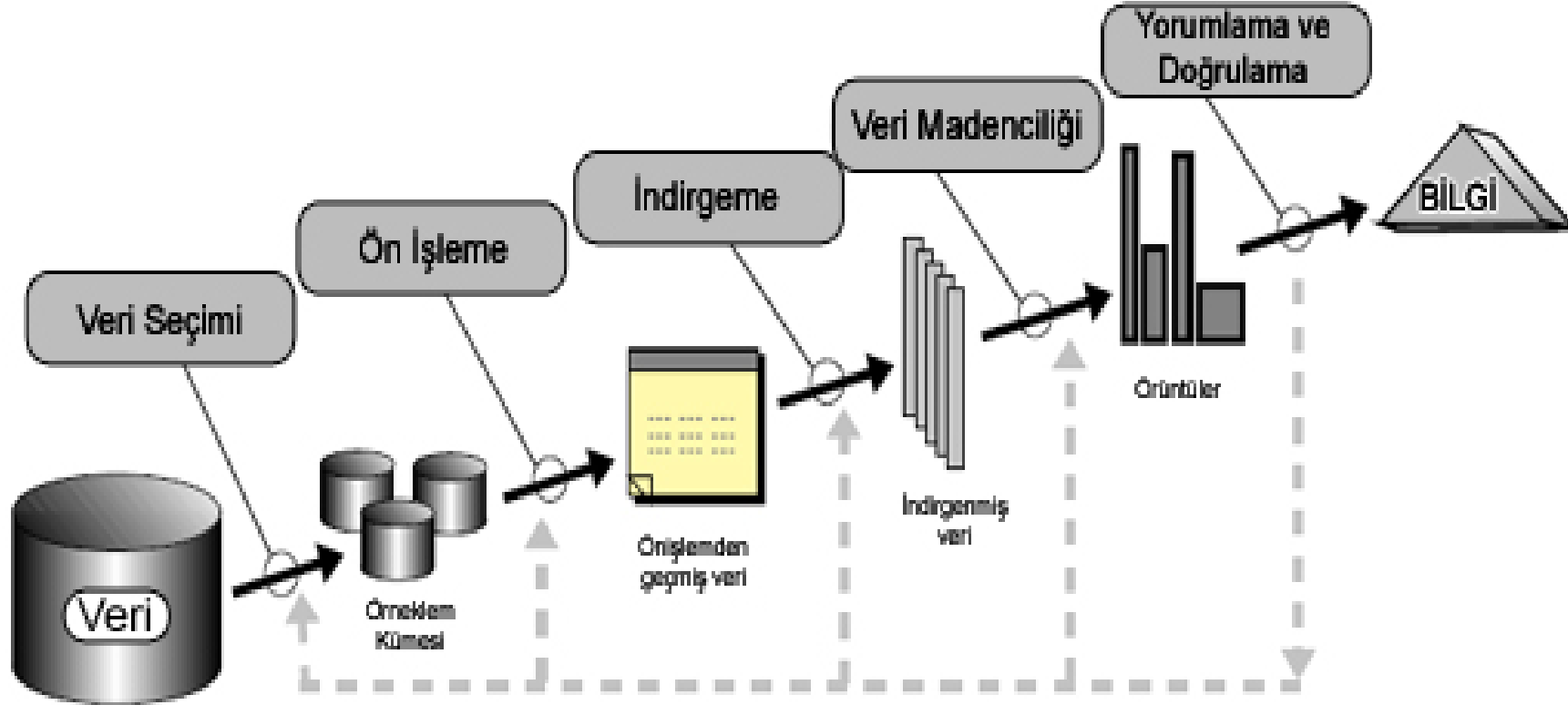


Veri Madenciliği Süreci

- Veri temizleme
- Veri bütünleştirme
- Veri indirgeme
- Veri dönüştürme
- Veri madenciliği algoritmasını uygulama
- Sonuçları sunma ve değerlendirme



Veri Madenciliği Süreci



Veri Madenciliği Süreci

Veri temizleme

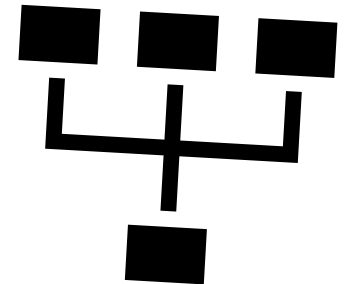
- Veritabanında yer alan tutarsız ve hatalı veriler gürültü olarak değerlendirilir. Bu gibi durumlarda verinin söz konusu sorunlardan temizlenmesi gerekir.
- Eksik değer içeren kayıtlar veri kümesinden atılabilir.
- Kayıp değerlerin yerine bir genel sabit kullanılabilir, fakat bütün değişkenlerde sabit değer kullanımı sorun yaratabilir.
- Değişkenin tüm verilerinin ortalaması alınarak eksik değer yerine bu ortalama değer kullanılabilir.
- Tüm veriler yerine sadece bir sınıfa ait verilerin ortalaması alınabilir.
- Verilere uygun bir tahmin yapılarak eksik değer yerine tahmini değer kullanılabilir.



Veri Madenciliği Süreci

Veri bütünleştirme

- Farklı veri tabanlarından yada veri kaynaklarından elde edilen verilerin birlikte değerlendirilmeye alınabilmesi için farklı türdeki verilerin tek türe dönüştürülmesi yani bütünleştirilmesi söz konusu olacaktır.
- Veri bütünleştirme ile farklı kaynaklardan gelen verilerin tutarsızlıkları giderilir.



Veri Madenciliği Süreci

Veri İndirgeme

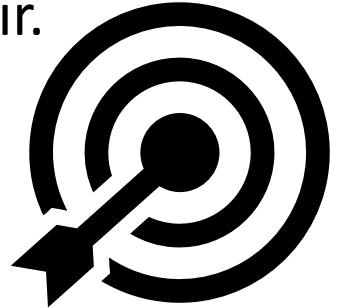
- Veri madenciliği uygulamalarında bazen çözümleme işlemi uzun süre alabilir. Eğer çözümlemeden elde edilecek sonucun değişmeyeceğine inanılıyorsa veri sayısı ya da değişkenlerin sayısı azaltılabilir.
- Veri indirgeme değişik boyutlarda yapılabilir;
 - a)Veriyi birleştirme veya veri küpü
 - b)Boyut indirgeme
 - c)Veri sıkıştırma
 - d)Örnekleme
 - e)Genelleme



Veri Madenciliği Süreci

Veri İndirgeme

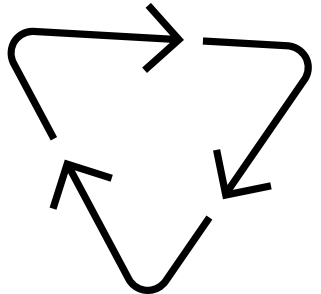
- **Veriyi indirgeme** aşamasında verileri çok boyutlu veri küpleri biçimine dönüştürmek söz konusu olabilir. Böylece çözümler sadece belirlenen boyutlara göre yapılır. Veriler arasında bir seçme işlemi yapılarak, gereksiz veriler veritabanından çıkarılır ve boyut azaltılması sağlanabilir.
- **Veri sıkıştırma** aşamasında, büyük veri kümelerinin sıkıştırılarak daha az yer işgal etmeleri sağlanır.
- **Örnekleme** aşamasında ise, büyük veri topluluğu yerine onu temsil eden daha küçük veri kümelerinin oluşturulması amaçlanır.
- **Genellemede** verilerin tek tek değil genel kavramlarla ifade edilmesi sağlanır.



Veri Madenciliği Süreci

Veri dönüştürme

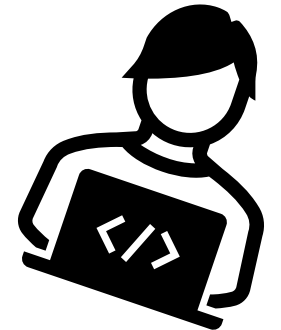
- Bazı durumlarda orijinal veriyi, veri madenciliği çözümlerine olduğu gibi katmak uygun olmayabilir.
- Değişkenlerin ortalama ve varyansları birbirinden önemli ölçüde farklı olduğu takdirde büyük ortalama ve varyansa sahip değişkenlerin diğerleri üzerindeki baskısı daha fazla olur ve onların önemli rollerini önemli ölçüde azaltır.
- Bu nedenle bir dönüşüm yöntemi uygulanarak söz konusu değişkenlerin normalleştirilmesi veya standartlaştırılması uygun bir yol olacaktır.



Veri Madenciliği Süreci

Veri madenciliği algoritmasını uygulama

- Veri madenciliği algoritmalarını uygulamadan önce, yukarıda sayılan ön işlemler ile veriyi hazır hale getirmek gerekir.
- İhtiyaç duyulan veri ön işleme yöntemleri uygulandıktan sonra amaca yönelik veri madenciliği algoritması ile veri madenciliği sürecine devam edilir.
- Temel Veri madenciliği yöntemleri: sınıflandırma, kümeleme ve birliktelik kuralları.



Veri Madenciliđi Süreci

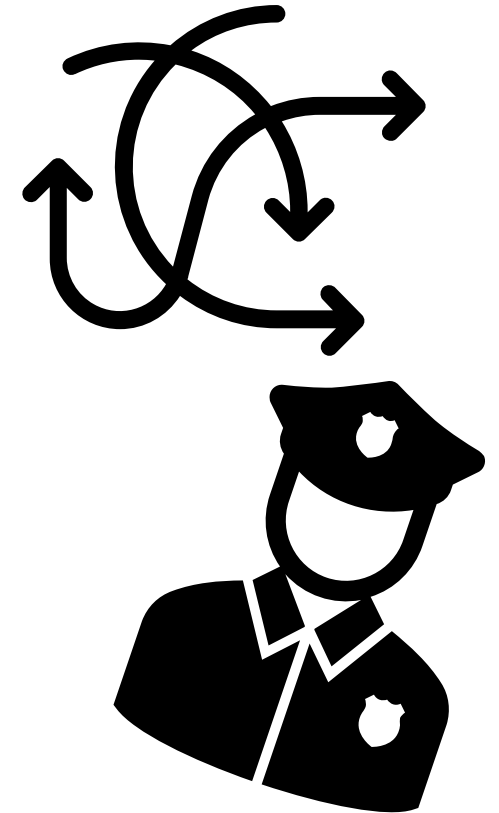
Sonuçları sunma ve deđerlendirme

- Veri madenciliđi algoritması veriler üzerinde uygulandıktan sonra, sonuçlar düzenlenerek ilgili yerlere sunulur.
- Sonuçlar çođu kez grafiklerle desteklenir.



Veri Madenciliğinin Zorlukları

- Teknik
 - ☐ Büyük Veri (tera-byte,peta-byte, ...)
 - ☐ Karmaşık, çoklu-ortamlı, yapısal olmayan veriler
 - ☐ Alan bilgisi ile entegrasyon
- İşletme
 - ☐ İyi uygulama alanları bulma zorluğu
- Toplumsal
 - ☐ Gizlilik & Güvenlik sorunları
 - ☐ Yasal problemler



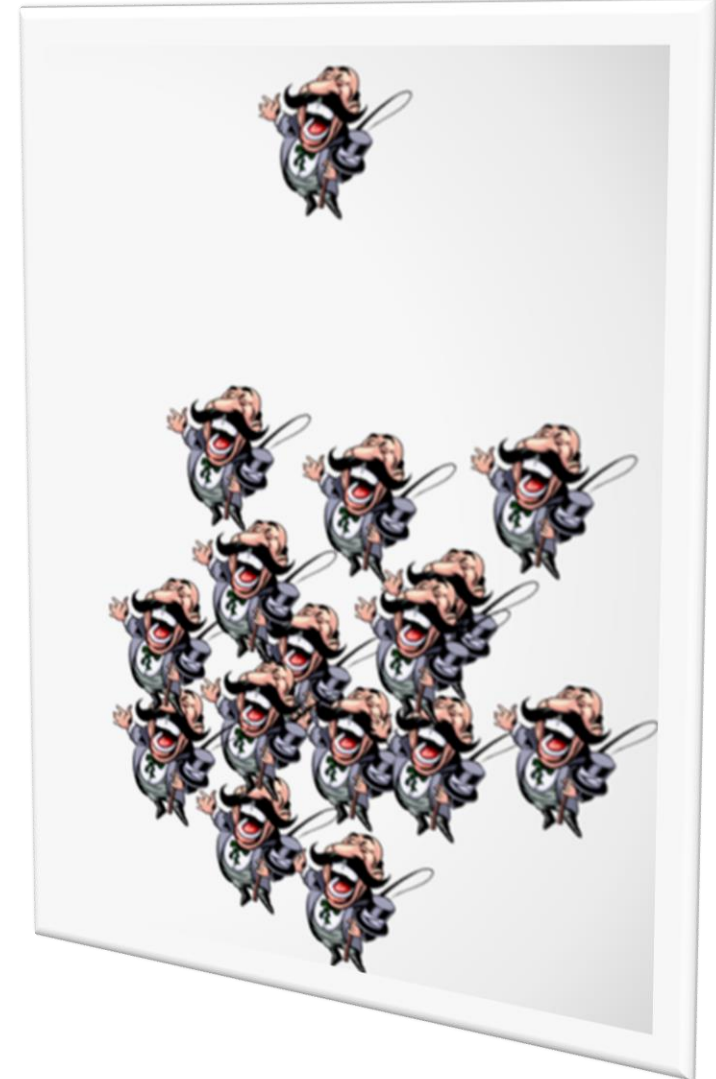
ÖZET

- Veri madenciliğinde amaç çok büyük miktardaki ham veriden değerli bilginin çıkarılmasıdır.
- Çıkarılan kuralların kalitesi öncelikle verinin kalitesine bağlıdır. Temiz ve güvenilir veri, veri madenciliğinin ön şartıdır.
- Bu alanda çalışan kişilerin alanlarında uzman olmaları gerekmektedir. Günümüzde her ne kadar daha az müdahale gerektiren programlar kullanılsa da verinin sisteme verilmesi ve sonuçların yorumlanması için uzman kişilere ihtiyaç vardır.
- Veri madenciliği tek aşamalı bir çalışma değildir, tekrarlıdır. Sistem ayarlanana dek birçok deneme gerektirebilir. Çalışma uzun olabilir. Buna çalışan ekibin ve yönetimin hazırlıklı olması, kısa vadede çok büyük beklentilere sahip olunmaması gerekir.

ÖZET

Verilerin konuşmasına müsaade edin...

Söyleyeceği çok şeyi olabilir...fakat sadece biraz
gürültülü olabilir.



Referanslar

Veri madenciliği Kavram ve Algoritmaları, Gökhan Silahtaroğlu, Papatya Yayıncılık

Veri Madenciliği Yöntemleri, Yalçın Özkan, Papatya Yayıncılık

DATA MINING: Concepts, Models, Methods & Algorithms, Second Edition, John Wiley, 2011, M. Kantardzic

Veri madenciliği ders notları, Dr. Burcu Çarklı YAVUZ, SAÜ

Veri madenciliği ders notları, Doç. Dr. Nilüfer YURTAY, SAÜ

Veri madenciliği ders notları, Dr. Tuğrul TAŞCI, SAÜ

Dr. Öğr. Üyesi Halil İbrahim CEBECİ Ders Notları, SAÜ

Database Systems: Design, Implementation, and Management; Carlos Coronel, Steven Morris, Peter Rob; Cengage Learning (Genel İçeriğin Hazırlanması)

<https://www.i-scoop.eu/big-data-action-value-context/data-information-content-knowledge-input-insight-action-value/>

<https://github.com/celalceken/DatabaseManagementSystems>

www.internetlivestats.com/twitter-statistics/

<https://www.free-powerpoint-templates-design.com/>

<https://gifer.com/en/gifs/computer>

<https://tenor.com/search/coding-gifs>

