

VERİ MADENCİLİĞİ

Dr. Ali Serhan KOYUNCUGİL

BÖLÜM 4 VERİ MADENCİLİĞİ İLE DİĞER ANALİTİK YÖNTEMLERİN KARŞILAŞTIRILMASI

Dr. Ali Serhan KOYUNCUGİL

BÖLÜM 4 VERİ MADENCİLİĞİ İLE DİĞER ANALİTİK YÖNTEMLERİN KARŞILAŞTIRILMASI

- 4.1. İstatistiksel Veri Analizi
- 4.2. Çevrimiçi Analitik İşleme (OLAP)
- 4.3. Yapısal Sorgu Dili (SQL) veya Veritabanı Sorgusu
- 4.4. İstatistiksel analiz ile veri madenciliğinin karşılaştırılması
- 4.5. Veri madenciliği ile OLAP'ın Kıyaslanması
- 4.6. Veri madenciliği ile veri sorgusunun kıyaslanması
- 4.7. Keşfedilecek Bilgi Tipine Göre Uygun Yöntem Belirleme

BÖLÜM 4 VERİ MADENCİLİĞİ İLE DİĞER ANALİTİK YÖNTEMLERİN KARŞILAŞTIRILMASI

Veri Madenciliği zaman zaman

- Büyük bir veri ambarı üzerinde SQL sorgusu,
- Herhangi bir sayıda veritabanı veya veri ambarları üzerinde SQL sorgusu,
- Çok boyutlu veritabanı analizi ,
- OLAP,
- Açıklayıcı veri analizi,
- İleri grafiksel görselleştirme,
- Veri ambarı üzerinde geleneksel istatistiksel işleme vb. ile **karıştırılmaktadır. Ancak HİÇBİRİSİ DEĞİLDİR.**

4.1. İSTATİSTİKSEL VERİ ANALİZİ

4.1. İSTATİSTİKSEL VERİ ANALİZİ

İstatistik, rasgele olaylar, süreçler, sistemler hakkında, modeller kurma, hipotezler kurma, hipotezlerin geçerliliğini sınamak, simülasyonlar ve tahmin yoluyla karar destek sağlamak olarak tanımlanabilir.

Ayrıca, bilinmeyen parametrenin bir fonksiyonuna da istatistik denir.

İstatistik analiz ise bir süreç olarak veri tanımlama, veri toplama, veri düzenleme, uygun yöntemlerle test etme veya modelleme ile sonuç çıkarımı olarak tanımlanabilir.

4.2. ÇEVİRİMİÇİ ANALİTİK İŞLEME (OLAP)

4.2. ÇEVİRİMİÇİ ANALİTİK İŞLEME (OLAP)

Veri tabanları (ambarları) üzerinde çok boyutlu tablolamayla, hızlı biçimde özet veriler sunma (miktar, ortalama, standart sapma vb.) sürecine Çevrimiçi Analitik İşleme denilmektedir.

OLAP işlemi çok boyutlu veritabanlarında veya çok boyutlandırılmış veritabanlarında gerçekleştirilmektedir.

4.3. YAPISAL SORU DİLİ (SQL) VEYA VERİTABANI SORGUSU

4.3. YAPISAL SORU DİLİ (SQL) VEYA VERİTABANI SORGUSU

SQL (Structured Query Language – Yapılandırılmış Soru Dili), ilişkisel veri tabanlarında kullanılan bir yazılım dilidir. Veri yüklemek, silmek, çağırmak, değiştirmek vb. işlemlerde kullanılır.

SQL ve benzeri sorgular ile veri tabanlarında yer alan verilerden istenilen raporlar, istenilen formatta elde edilebilir.

4.4. İSTATİSTİKSEL ANALİZ İLE VERİ MADENCİLİĞİNİN KARŞILAŞTIRILMASI

4.4. İSTATİSTİKSEL ANALİZ İLE VERİ MADENCİLİĞİNİN KARŞILAŞTIRILMASI

- | İstatistiksel Analiz | Veri Madenciliği |
|---|--|
| • İstatistikçiler genellikle bir hipotez ile başlarlar. | • Veri madenciliği hipoteze gerek duymaz. |
| • İstatistikçiler hipotezlerini eşleştirmek için kendi eşitliklerini geliştirmek zorundadırlar. | • Veri madenciliği algoritmaları eşitlikleri otomatik olarak geliştirir. |
| • İstatistiksel analizler sadece sayısal verileri kullanır. | • Veri madenciliği farklı tiplerde data kullanır (örneğin metin, ses) sadece sayısal veriyi değil. |

4.4. İSTATİSTİKSEL ANALİZ İLE VERİ MADENCİLİĞİNİN KARŞILAŞTIRILMASI

İstatistiksel Analiz

Veri Madenciliği

- İstatistikçiler kirli veriyi analizleri sırasında bulur ve dayanır. filtre ederler.
- İstatistikçiler kendi sonuçlarını yorumlar ve bu sonuçları yöneticilere iletirler.
- Veri madenciliği temiz veriyi yorumlamak kolay değildir. Veri madenciliği sonuçlarını analiz etmede ve yorumlamada ve bulguları yöneticilere iletmede mutlaka İSTATİSTİKÇİYE ihtiyaç duyulmaktadır.

4.5. VERİ MADENCİLİĞİ İLE OLAP'IN KİYASLANMASI

4.5. VERİ MADENCİLİĞİ İLE OLAP'IN KİYASLANMASI

Büyük veri tabanlarında basit ilişkiler keşfedilmek isteniyorsa OLAP kullanılmalıdır.

4.6. VERİ MADENCİLİĞİ İLE VERİ SORGUSUNUN KİYASLANMASI

4.6. VERİ MADENCİLİĞİ İLE VERİ SORGUSUNUN KİYASLANMASI

Hemen hemen ne arandığı biliniyor
ve
büyük veri tabanı ile çalışmak isteniyorsa
veri sorgusu kullanılmalıdır.

4.7. KEŞFEDİLECEK BİLGİ TİPİNE GÖRE UYGUN YÖNTEM BELİRLEME

4.7. KEŞFEDİLECEK BİLGİ TİPİNE GÖRE UYGUN YÖNTEM BELİRLEME

Veri içerisinde açıkça gözlenemeyen örüntü ve ilişkiler bulunmak isteniyorsa veri madenciliği kullanılmalıdır.

4.7. KEŞFEDİLECEK BİLGİ TİPİNE GÖRE UYGUN YÖNTEM BELİRLEME

SQL, OLAP ve veri madenciliği kullanımını, keşfedilmek istenen bilgi tipine göre sınıflarsak:

Sığ Bilgi: Seçilen kayıtlara ait ortalama ve toplam değer gibi özet bilgiler için kayıt seçmek yeterlidir ki SQL bunu yapabilir.

Çok boyutlu bilgi: Farklı özelliklerin, ortaya çıkma sıklığı hakkında bilgi. Veri küpü üzerinde OLAP bunu yapabilir.

Gizli bilgi: Önceden tahmin edilemeyen örüntü ve ilişkiler veri madenciliği için başlangıç olabilir.

Derin bilgi: Sadece önsel teknik veya meta-bilginin kullanımıyla keşfedilebilecek gizli örüntüler ve ilişkiler hakkında bilgi. Bu konu veri madenciliğinin araştırma sınırları içerisindedir.