

BAŞKENT ÜNİVERSİTESİ
TİCARİ BİLİMLER FAKÜLTESİ
Yönetim Bilişim Sistemleri Bölümü
TBF 491

VERİ MADENCİLİĞİ

Dr. Ali Serhan KOYUNCUGİL

BAŞKENT ÜNİVERSİTESİ
TİCARİ BİLİMLER FAKÜLTESİ
Yönetim Bilişim Sistemleri Bölümü
TBF 491

BÖLÜM 3 VERİ MADENCİLİĞİNİN BİLEŞENLERİ

Dr. Ali Serhan KOYUNCUGİL

BÖLÜM 3 VERİ MADENCİLİĞİNİN BİLEŞENLERİ

- 3.1. (Çok Değişkenli) İstatistiksel Veri Analizi
- 3.2. Makine Öğrenimi
- 3.3. Örüntü Tanıma
- 3.4. Yapay Zeka
- 3.5. Veri Tabanları
- 3.6. Uzman Sistemler
- 3.7. Veri Görselleştirme
- 3.8. Yüksek Performanslı (Hızlı) Hesaplama

3.1. (ÇOK DEĞİŞKENLİ) İSTATİSTİKSEL VERİ ANALİZİ

3.1. (ÇOK DEĞİŞKENLİ) İSTATİSTİKSEL VERİ ANALİZİ

Veri kümesinin içerdği değişkenler arasındaki yapısal ilişkiyi, ilişkinin büyüklüğünü ve yönünü, değişim miktarını ve gelecekteki olası değerlerini belirlemek üzere gerçekleştirilen, istatistiki yöntemlere dayalı veri analiz işlemine istatistiksel veri analizi denir.

İstatistiksel veri analizi veri toplama, veri seçimi, veri temizleme, veri analizi, tahmin ve yorum başlıkları altında incelenebilir.

Veri kümeleri özünde çok boyutlu ve değişkenlidir. Ancak, zaman zaman tek değişken ve iki değişkenli durumlar için de analiz ihtiyacı olabilir. İki'den fazla değişken etüdü ise çok değişkenli istatistiksel veri analizi kapsamında incelenir.

3.2. MAKİNE ÖĞRENİMİ

3.2. MAKİNE ÖĞRENİMİ

Bir problemi çözmek için örnek veri veya geçmiş tecrübeleri kullanmak üzere bilgisayarları programlamaya Makine Öğrenimi denir.

3.3. ÖRÜNTÜ TANIMA

3.3. ÖRÜNTÜ TANIMA

İnsanoğlu

- Yüzleri tanıma,
- El yazısını okuma,
- Konuşmaları anlama,
- Sesleri algılama ve ayırma

gibi pek çok farklı işlevi yerine getirecek yetenektedir.

Yukarıda bahsedilen ve benzeri işlevlerin makineler tarafından da yerine getirilmesi yönündeki çalışmalar örüntü tanınmanın ortaya çıkmasına yol açmıştır.

3.3. ÖRÜNTÜ TANIMA

- El yazısı,
 - İnsan yüzü,
 - Müzik sesi,
 - Coğrafi uydu görüntüleri,
- gibi kesinlik içermeyen ve belirli belirsiz tanımlanabilen olgulara **örüntü** denir.

3.3. ÖRÜNTÜ TANIMA

Söz konusu örüntülerin makineler tarafından,

- Algılanması,
- İşlenmesi,
- Ayırt edilmesi,
- Sınıflandırılması,
- Eşleştirilmesi

gibi işlemlerin, doğru karar verecek biçimde gerçekleştirilmesine yönelik çalışmalara **örüntü tanıma** denir.

3.4. YAPAY ZEKA

3.4. YAPAY ZEKA

Yapay zeka (YZ - Artificial Intelligence - AI) yapay bir varlığın zeka göstermesi olarak tanımlanmaktadır. Bu tür sistemler genelde bilgisayarlar olarak varsayılmaktadır.

Bilgisayar bilimlerinin temel konularından olan YZ makinelerin akıllı davranış göstermesi, öğrenme ve adaptasyon konularıyla ilgilenmektedir.

3.4. YAPAY ZEKA

YZ konusundaki araştırmalar, üretim makinelerinden, otomasyona dayalı görevlere kadar zeka gerektiren konuları dikkate almaktadır. Örnekler, kontrol, planlama, zamanlama gibi tanısız ve tüketici cevaplarına yanıt verecek biçimde, el yazısı, konuşma ve yüz tanıma gibi gerçek dünya problemlerine çözüm bulacak niteliktedir. YZ sistemleri, ekonomi, tıp, mühendislik ve askeri alanlar kadar, satranç gibi klasik strateji oyunlarından, diğer oyunlara kadar ev bilgisayarları yazılım uygulamalarında da kullanılmaktadır.

3.5. VERİ TABANLARI

3.5. VERİ TABANLARI

Verileri, erişim ve kullanım kolaylığı sağlayacak biçimde belirli bir mantık içerisinde depolayan yapılara veritabanı denilmektedir.

Veritabanları yapısal (structured) yani tanım ve standartları belirli verileri, aynı mantıkta depolayacak biçimde çalışırlar.

Belirli tanım, standart ve metodoloji altında depolanmayan veriler ise yapısal olmayan (unstructured) veri olarak adlandırılırlar.

3.5. VERİ TABANLARI

Veritabanları genel anlamda,

- Hiyerarşik,
 - İlişkisel,
 - Nesne yönelimli
- başlıkları altında çeşitlendirilebilir.

3.6. UZMAN SİSTEMLER

3.6. UZMAN SİSTEMLER

Uzman bilgisini, akıllı sistemler aracılığıyla sunan sistemlere Uzman Sistem denilmektedir.

Yapay zeka ile uzman sistem benzeşmesi oldukça fazla olmasına karşın aralarındaki nüans,

- Yapay zeka, insan gibi düşünüp karar vermek,
- Uzman sistem ise çalışılan konunun uzman bilgisine göre karar vermek üzere tasarlanmıştır.

3.7. VERİ GÖRSELLEŞTİRME

3.7. VERİ GÖRSELLEŞTİRME

Veriyi doğru analiz etmek, veriyi doğru anlamak ve tanımaktan geçmektedir. Betimsel veri madenciliği/istatistiki yöntemler veri hakkında sayısal bilgi verseler de, sadece sayısal değerlere bakarak karar vermek zaman zaman yanıltıcı olabilmektedir. Ayrıca, yapılandırılan modelin tahmin sonuçlarını değerlendirmek, gerçekleştirmelerle karşılaştırmak gibi durumlarda görsel destek algılama ve anlamada kolaylık sağlamaktadır.

3.7. VERİ GÖRSELLEŞTİRME

Veri görselleştirme, verilerin grafik yöntemler aracılığıyla sunumu olarak ifade edilebilir ve karar verme sürecinde görsel katkı sağladığı ifade edilebilir.

Çubuk grafik, pasta grafik, saçılım grafiği, histogram gibi betimsel görsel yöntemler bu kapsamda incelenebileceği gibi, hiyerarşik kümeleme analizindeki dendogram veya karar ağaçları yöntemlerindeki ağaç grafikleri gibi tahminsel grafikler de bu kapsamda ele alınabilir.

3.8. YÜKSEK PERFORMANSLI (HIZLI) HESAPLAMA

3.8. YÜKSEK PERFORMANSLI (HIZLI) HESAPLAMA

Bu terim genellikle, bilimsel araştırmalar veya bilgisayar bilimleri ile ilişkili olarak ele alınmaktadır. İlişkili terim olan yüksek performanslı teknik hesaplama, küme (cluster) tabanlı hesaplamanın mühendislik uygulamalarına atıfta bulunmaktadır.

Mantık olarak, çok yüksek konfigürasyonlu olmayan bir bilgisayarla çok uzun zaman alacak bir hesaplamanın, çok yüksek konfigürasyonlu bir veya küme mantığıyla çalışan birden fazla makinede (fiziki olarak aynı mekanda olsun veya olmasın) çok kısa sürede hesaplanmasıdır. Yüksek performanslı hesaplama, zaman zaman süper hesaplama ile eş anlamlı olarak anılmaktadır. Ancak, süper hesaplama, yüksek performanslı hesaplamanın daha güçlü bir alt kümesidir.