VERİ MADENCİLİĞİ (DATA MİNİNG)

Dr. Öğr. Üyesi Alper Talha Karadeniz

Ders İçeriği

01

Veri Madenciliğine Giriş ve Uygulama Alanları 02

Veri Tipleri , Veri Yapıları ve Temel Kavramlar

03

Veri Temizleme ve Ön İşleme Teknikleri 04

Veri Görselleştirme ve Keşifsel Veri Analizi (EDA) 05

Temel Benzerlik ve Uzaklık Ölçütleri 06

Kümeleme Yöntemleri

07

Sınıflandırma Algoritmaları 08

Regresyon Modelleri 09

İlişki Kuralları Madenciliği 10

Metin Madenciliği ve Duygu Analizi 11

Büyük Veri ve Gerçek Zamanlı Madencilik 12

Güncel Yaklaşımlar ve Uygulama Projesi Sunumları

Hafta 1

01

Veri Madenciliği Tanımı ve Kapsamı 03

Veri Bilimi Makine Öğrenimi ve istatistik ile ilişkisi

05

Uygulama

02 Bilgi Keşfi (KDD) Süreci

04 Modern Kullanım Alanları

Veri Madenciliği Nedir?

Veri madenciliği, büyük ve karmaşık veri kümeleri içerisinden daha önce fark edilmemiş, anlamlı ve kullanılabilir bilgileri keşfetme sürecidir. Bu süreç, istatistik, makine öğrenmesi ve veri tabanı teknolojilerini bir araya getirerek, verilerdeki örüntüleri, ilişkileri ve eğilimleri ortaya çıkarır. Genellikle "Veri Tabanlarında Bilgi Keşfi" (KDD) sürecinin bir parçası olarak değerlendirilir ve veri temizleme, bütünleştirme, seçme, dönüştürme, madencilik, örüntü değerlendirme ve bilgi sunumu gibi adımları içerir. Veri madenciliği sayesinde işletmeler; müşteri davranışlarını analiz edebilir, riskleri öngörebilir, ve stratejik kararlar alabilir. Bu yönüyle veri madenciliği, sadece teknik bir analiz yöntemi değil, aynı zamanda karar destek sistemlerinin temel bileşenlerinden biridir.

IDC'ye göre, dijital evrende üretilen toplam veri miktarı 2025 yılı itibarıyla yaklaşık 1.2 yottabyte, yani 1,208,925,819,614,630,000,000,000 byte seviyesine ulaşıyor. Bu büyüklük, veri madenciliği süreçlerinin sadece teknik değil aynı zamanda küresel ölçekte stratejik bir öncelik haline geldiğini gösteriyor

WEB

1.13 milyar web sitesi mevcut (2025 itibarıyla=

Çevrimiçi Sosyal Ağlar

İnstagram 2 milyar aylık aktif kullanıcı



Öneri Sistemleri

Netfilx'te 301 milyon global abone

Wikipedia

İngilizce 7.031.896 madde tüm dillerde 60 milyon+ madde

Samsun Ünive Yazılım Mühe

Neden Veri Madenciliği Yapıyoruz?

- Gizli Bilgileri Ortaya Çıkarmak Büyük veri yığınlarında saklı kalmış ve önceden fark edilmemiş değerli bilgileri keşfetmek için yapılır.
- Maliyetleri Azaltmak ve Verimliliği Artırmak Gereksiz işlemleri önleyerek kurumların kaynak kullanımını daha verimli hale getirir.
- Karmaşık Veri Yapılarını Anlamak Yapısal olmayan veri kaynaklarında (metin, görüntü, sensör verisi gibi) örüntü bulmak amacıyla uygulanır.
- Geleceği Tahmin Etmek (Öngörü Analizi) Müşteri davranışları, satışlar ya da risk düzeyleri gibi konularda geleceğe yönelik tahminlerde bulunmak için.
- Rekabet Baskısına Yanıt Vermek Hızla değişen pazarlarda rakiplerin önüne geçebilmek için veri madenciliğiyle daha hızlı, doğru ve kişiselleştirilmiş kararlar alınır. Bu sayede kurumlar, müşteri beklentilerine daha iyi yanıt verir ve pazarda avantaj sağlar.





Bilgi KeŞif Süreci ve Aşamaları

Ham verinin toplanması, temizlenmesi, dönüştürülmesi ve madencilik yöntemleriyle işlenmesi sonucu bilgiye dönüştürüldüğü bütün sürece KDD denir

Amaca yönelik olarak veri Veri Seçimi kaynaklarından ilgili veriler seçilir. Eksik, hatalı veya tutarsız veriler Veri Ön İşleme düzeltilir ya da çıkarılır. Gürültü (noise) azaltılır. Veriler analiz için uygun forma getirilir. Veri Dönüştürme Normalleştirme, öznitelik seçimi, boyut indirgeme yapılabilir. Temel bilgi çıkarma adımıdır Veri Madenciliği Belirli algoritmalar kullanılarak kalıplar, ilişkiler, modeller bulunur.

Samsun Üniversitesi Yazılım Mühendisliği Bölümü Yorumlama ve Değerlendirme Elde edilen bilgilerin doğruluğu, anlamlılığı ve faydası değerlendirilir

Veri Madenciliği & İstatistik

- Amaçları benzerdir.
- Farklı yöntemler kullanılır.
- Veri madenciliğinde daha fazla hipotez araştırılır.
- Veri madenciliği daha büyük verilerle uğraşır.
- Veri madenciliği **keşifsel** (**exploratory**) yaklaşımla çalışır, daha önce bilinmeyen yapıları ortaya çıkarır.
- Veri madenciliği uygulama istatistik teori temllidir.
- Veri madenciliği KDD sürecinin bir parçası iken istatistik bilimsel araştırmaların temel analiz aracıdır.



Veri Madenciliği & Makine Öğrenimi

- Veri madenciliği, makine öğrenimi algoritmalarını kullanır. Karar ağaçları, kümeleme, sınıflandırma gibi pek çok veri madenciliği yöntemi, makine öğrenimi temellidir.
- Veri madenciliği, bilgi keşfine; makine öğrenimi ise öğrenmeye odaklanır. Ancak her ikisi de veriden anlam çıkarmayı hedefler.
- Her iki alan da büyük veriyle çalışmak için uygundur.
- Makine öğrenimi modelleri, veri madenciliği sonuçlarının otomatikleştirilmesini sağlar. Böylece keşfedilen bilgilerin gerçek zamanlı tahmin sistemlerine dönüştürülmesi mümkün olur.
- İki alan da veri biliminin temel bileşenlerindendir. Veri biliminde hem veri madenciliğiyle keşif yapılır, hem de makine öğrenimiyle tahminsel modeller geliştirilir.



Modern Kullanım Alanları

Pazarlama ve Müşteri İlişkileri Yönetimi (CRM)

- Müşteri segmentasyonu
- Hedefli kampanyalar oluşturma
- Müşteri kaybını (churn) tahmin etme

Finans ve Bankacılık

- Kredi risk analizi
- Dolandırıcılık (fraud) tespiti
- Müşteri davranış analizi

Sağlık Sektörü

- Hastalık teşhisi ve risk tahmini
- Hasta verilerinden tedavi önerileri
- Tıbbi görüntüleme verilerinin analizi

E-ticaret ve Öneri Sistemleri

- Kişiselleştirilmiş ürün önerileri
- Satın alma alışkanlıklarının analizi
- Sepet analizi (market basket analysis)

Modern Kullanım Alanları

Güvenlik ve Savunma

- Şüpheli faaliyetlerin tespiti
- Saldırı modelleme ve analiz
- Siber güvenlikte anomali tespiti

Üretim ve Sanayi (Endüstri 4.0)

- Arıza tahmini (predictive maintenance)
- Kalite kontrol süreçlerinin optimizasyonu
- Üretim verimliliği analizi

Eğitim

- Öğrenci başarı tahmini
- Öğrenme analitiği
- Öğrenci davranış analizi (e-öğrenme platformlarında)

Tarım ve Çevre

- Ürün verimliliği tahmini
- İklim ve toprak verisi analizi
- Doğal afet erken uyarı sistemleri

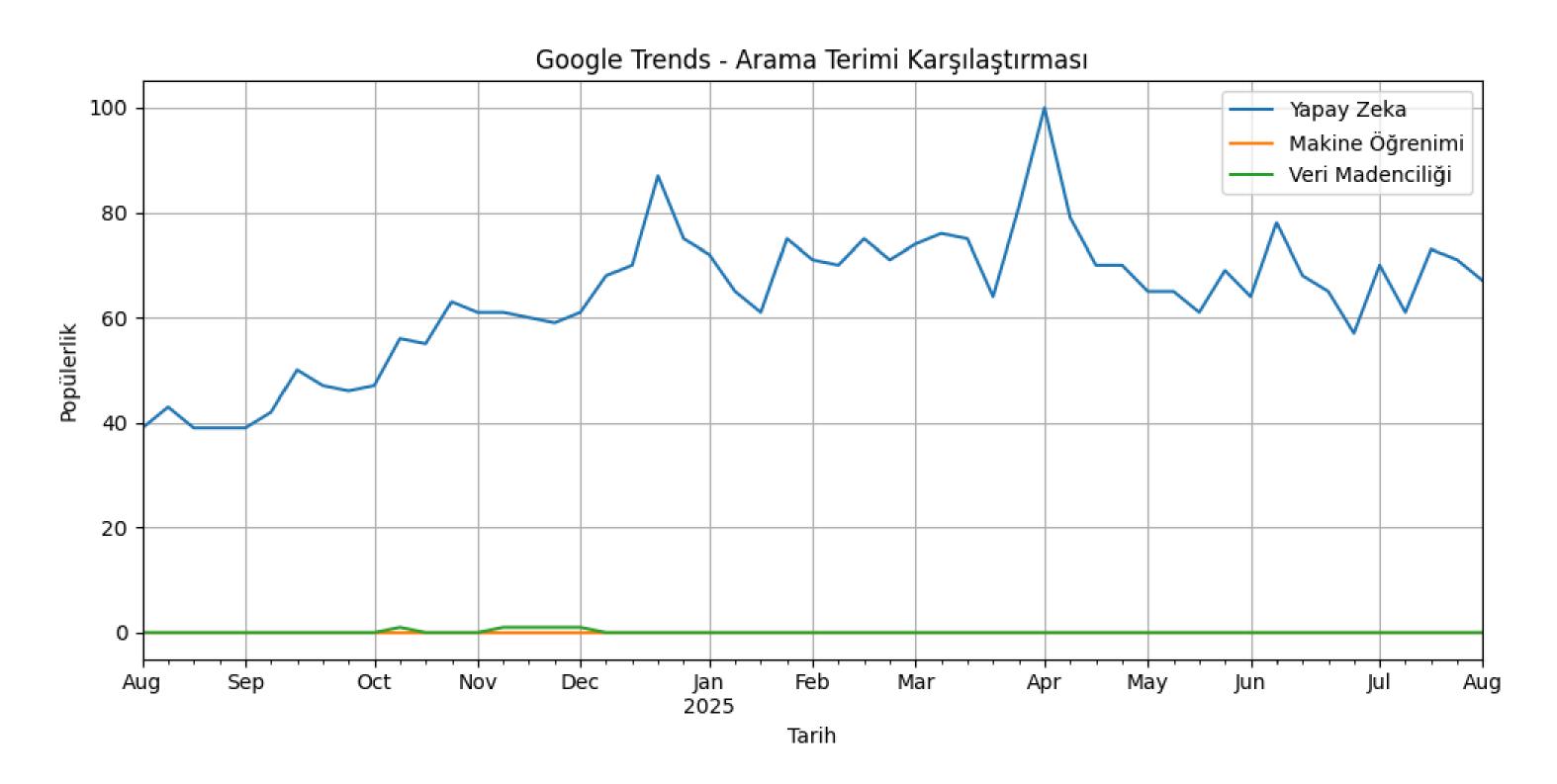
UYGULAMA

Google Trends

```
from pytrends.request import TrendReq
    import pandas as pd
    import matplotlib.pyplot as plt
    # Google Trends API'ye bağlan
    pytrends = TrendReq(hl='tr-TR', tz=360)
    # Arama terimlerini tanımla
    arananlar = ["Yapay Zeka", "Makine Öğrenimi", "Veri Madenciliği"]
10
    # Google Trends'ten verileri çek
    pytrends.build_payload(arananlar, timeframe='today 12-m')
    # Verileri al
    veri = pytrends.interest_over_time()
16
    # Grafikle göster
    veri[arananlar].plot(figsize=(10,5))
    plt.title("Google Trends - Arama Terimi Karşılaştırması")
    plt.xlabel("Tarih")
    plt.ylabel("Popülerlik")
    plt.grid(True)
    plt.tight_layout()
    plt.savefig("google_trends_grafik.png") # Grafik görseli olarak kaydet
    plt.show()
```

Google Trends, kullanıcıların belirli anahtar kelimeler hakkında yaptıkları arama sorgularının zaman içindeki popülerliğini gösteren ücretsiz bir analiz aracıdır. Belirli bir konunun ne zaman, nerede ve ne ölçüde ilgi gördüğünü analiz etmeye olanak tanır. . Bu veriler, kullanıcı eğilimlerini, sezonluk ilgi değişimlerini veya ani popülerlik artışlarını tespit etmek için sıkça kullanılır. Veri madenciliği ve desen keşfi çalışmaları için değerli bir veri kaynağıdır.

Uygulama: Google Trends



VERİ MADENCİLİĞİ (DATA MİNİNG)

Dr. Öğr. Üyesi Alper Talha Karadeniz