

VERİ MADENCİLİĞİ DERSİ

SUNU-1

Doç.Dr.Mehmet Akif ŞAHMAN

Ders Bilgileri

► Kaynaklar

- Kavram ve Algoritmalarıyla Temel Veri Madenciliği, Dr.Gökhan SİLAHTAROĞLU (Papatya Yayın Evi)
- İTÜ Veri Madenciliği Ders Notları, Şule Gündüz Öğüdücü,
- Veri Madenciliği Yöntemleri, Yalçın Özkan,
- Veri Madenciliği (Kavram ve Teknikler), Aysan Şentürk.
- İTÜ Veri Madenciliği Ders Notları, Kadriye ERGÜN,

► Ders Başarı Notu

- Vize (%40)
- Final (%60)



Veri Madenciliğine Giriş

- ▶ İçinde yaşadığımız bilişim çağında elektronik ortamda mevcut verinin hızlı artışı ve bilginin fazlalaşması sebebiyle öncelikle, genelde Veri Tabanlarında **Bilgi Keşfi** olarak adlandırılan yeni bir paradigma ortaya çıkmıştır. Daha yaygın bir kullanımla bu alana **Veri Madenciliği** denilmektedir.
- ▶ Büyük miktarda veri içinden, gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların aranmasıdır. Daha önceden bilinmeyen, geçerli ve uygulanabilir bilgilerin geniş veritabanlarından elde edilmesi ve bu bilgilerin işletme kararları verilirken kullanılmasıdır. Büyük ölçekli veriler arasından değeri olan bir bilgiyi elde etme işidir. *Yapısal* veritabanlarında depolanmış verilerden geçerli, yeni, potansiyel olarak yararlı ve nihayetinde anlaşılabilir örüntülerin tanımlanması işlemidir.



Veri Madenciliğine Giriş

- ▶ Bu tanımlamalardan da anlaşıldığı üzere veriler arasındaki ilişkileri ortaya koymak ve gerektiğinde ileriye yönelik tahminlerde bulunmak veri madenciliği çalışmaları sayesinde mümkün olmaktadır. Bunun anlamı, veri madenciliği bir kurumda üretilen tüm verilerin belirli yöntemler kullanarak var olan ya da gelecekte ortaya çıkabilecek gizli bilgiyi ortaya çıkarma süreci olarak değerlendirilmesidir. Bu açıdan bakıldığında veri madenciliği işinin kurumların Karar Destek Sistemleri için önemli bir yere sahip olduğu söylenebilir.
- ▶ Veri madenciliği çalışmaları, *sınıflandırma, ilişki kurma, kümeleme, regresyon, veri özetleme, değişikliklerin analizi, sapmaların tespiti* gibi belirli sayıda teknik yaklaşımları içerir.



Veri Madenciliğine Giriş

► Özet olarak

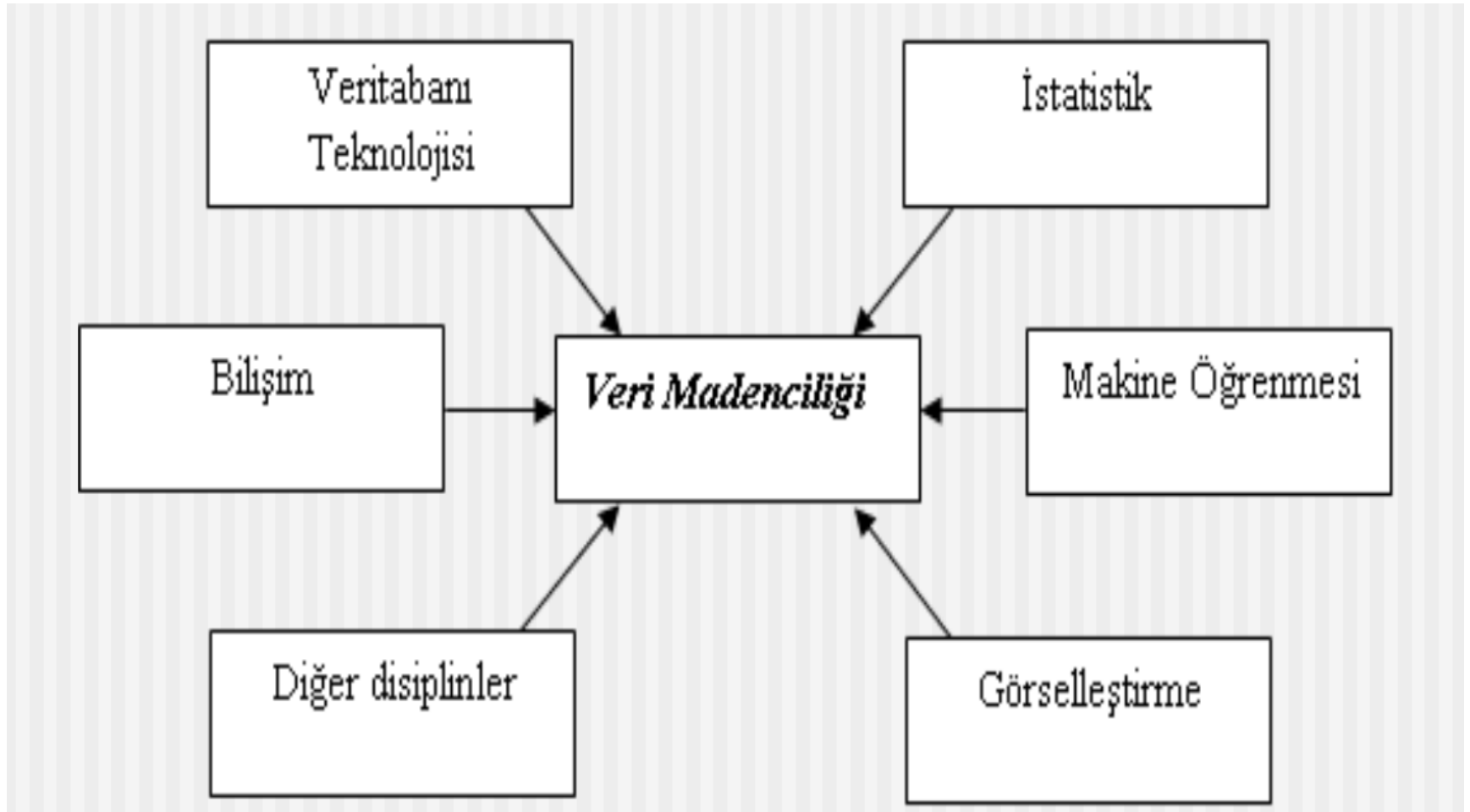
- Veri madenciliği, büyük veri setlerinden anlamlı bilgiler çıkarma sürecidir.
- Gelecekle ilgili tahminler yapmak, örüntüler keşfetmek ve karar destek sistemlerini güçlendirmek için kullanılır.

Örnek: Müşteri davranışlarını analiz ederek pazarlama stratejileri geliştirmek.

Tanım: "Büyük veri setlerinden, daha önce bilinmeyen, geçerli ve uygulanabilir bilgilerin elde edilmesi işlemidir."



Veri Madenciliğine Giriş



Veri Madenciliği Temel Kavramlar

- ▶ **Veri (Data)** : Kavramsal anlamda veri, kayıt altına alınmış her türlü olay, durum, fikirdir. Bu anlamıyla değerlendirildiğinde çevremizdeki her nesne bir veri olarak algılanabilir. (örneğin, sensör verileri)
 - ▶ **Enformasyon(Information)** : Verilerin ilişkilendirilmiş, düzenlenmiş, anlamlandırılmış, işlenmiş halidir. Bu haliyle enformasyon, potansiyel olarak içinde bilgi barından bir veri halindedir. (örneğin, sıcaklık ölçümlerinin grafiği)
 - ▶ **Bilgi (Knowledge)** : İşlenen verinin, veriyi değerlendirecek olan karar vericinin anlamlandırmasıdır. Dolayısıyla bireyin algılama yeteneği, yaratıcılık, deneyim gibi kişisel nitelikleri de bu süreci doğrudan etkilemektedir. (örneğin, sıcaklık artışının ekipman arızasına yol açabileceği bilgisi)
 - ▶ **Bilgelik (Wisdom)**: Bilgilerin bir araya getirilerek bir sentez oluşturulmasıdır. Yetenek, tecrübe gibi kişisel nitelikler birer bilgelik elemanıdır. (örneğin, ekipman bakım planlaması)
-



Veri Madenciliği Temel Kavramlar

**VERİ, ENFORMASYON,
BİLGİ VE BİLGELİĞİ BİR
OLAY ÜZERİNDE
ÖRNEKLENDİRİN.**



Veri Madenciliği Uygulama Alanları

Bilim	İş Hayatı	Web	Devlet
<ul style="list-style-type: none">• Astronomi• Biyoinformatik• İlaç keşfi	<ul style="list-style-type: none">• Reklam• CRM (Müşteri İlişkileri Yönetimi) ve Müşteri Modelleme• E-ticaret• Yatırım değerlendirme ve karşılaştırma• Sağlık• Üretim• Spor/eğlence• Telekom (telefon ve iletişim)• Hedef pazarlama	<ul style="list-style-type: none">• Metin Madenciliği (haber grubu, email, dokümanlar)• Web analizi• Arama motorları	<ul style="list-style-type: none">• Terörle Mücadele• Kanun Yaptırımı• Vergi Kaçakçılarının Profiline Çıkarılması



Veri Madenciliği Uygulama Alanları

- ▶ Veritabanı analizi ve karar verme desteği
 - ▶ Pazar araştırması
 - ▶ Hedef Pazar, müşteriler arası benzerliklerin saptanması, sepet analizi, çapraz pazar incelemesi
 - ▶ Risk analizi
 - ▶ Kalite kontrolü, rekabet analizi, öngörü
 - ▶ Sahtekarlıkların saptanması
- ▶ Diğer Uygulamalar
 - ▶ Belgeler arası benzerlik (haber kümeleri, e-posta)
 - ▶ Sorgulama sonuçları



Veri Madenciliğinin Elektrik-Elektronik Mühendisliğindeki Uygulamaları

- **Enerji Yönetimi:** Enerji tüketimi tahmini, akıllı şebekeler.
- **Sensör Verileri Analizi:** IoT cihazlarından gelen verilerin analizi.
- **Arıza Tespiti:** Makine öğrenmesi ile ekipman arızalarının önceden tahmin edilmesi.
- **Sinyal İşleme:** Gürültüden arındırma, sinyal sınıflandırma.
- **Üretim Optimizasyonu:** Üretim hatlarında verimliliği artırma.



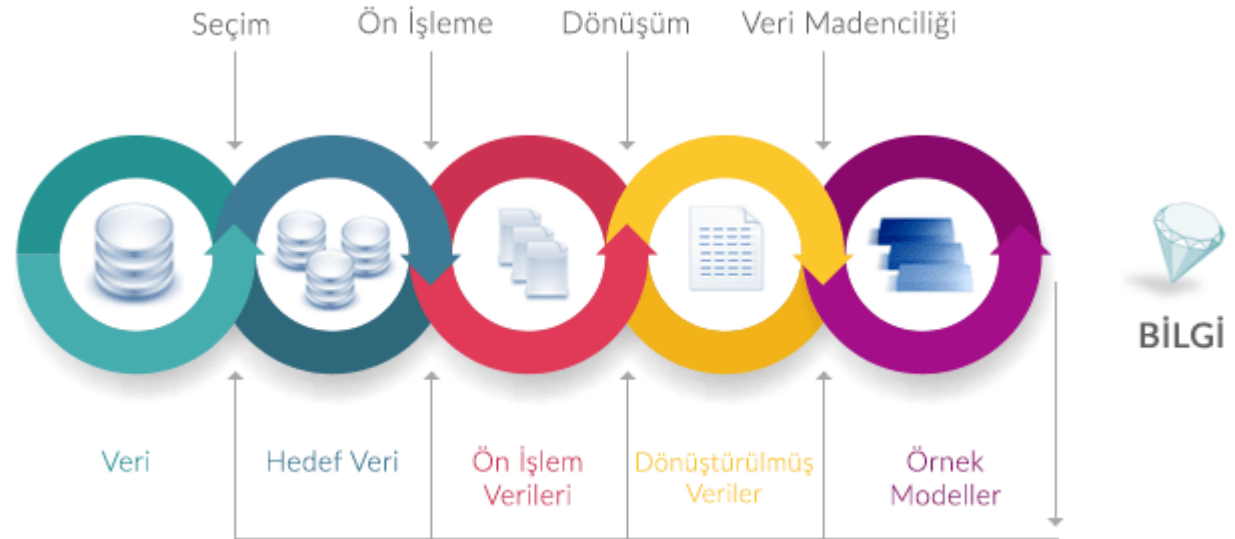
Veri Madenciliği Süreci

1. **Veri Toplama:** Ham verilerin toplanması (örneğin, sensörlerden, veritabanlarından).
2. **Veri Ön İşleme:** Veri temizleme, eksik verilerin tamamlanması, normalizasyon.
3. **Modelleme:** Algoritmaların uygulanması (sınıflandırma, kümeleme, regresyon).
4. **Değerlendirme:** Model performansının ölçülmesi ve iyileştirilmesi.
5. **Bilgi Keşfi:** Elde edilen bilgilerin raporlanması ve karar verme sürecine entegrasyonu.



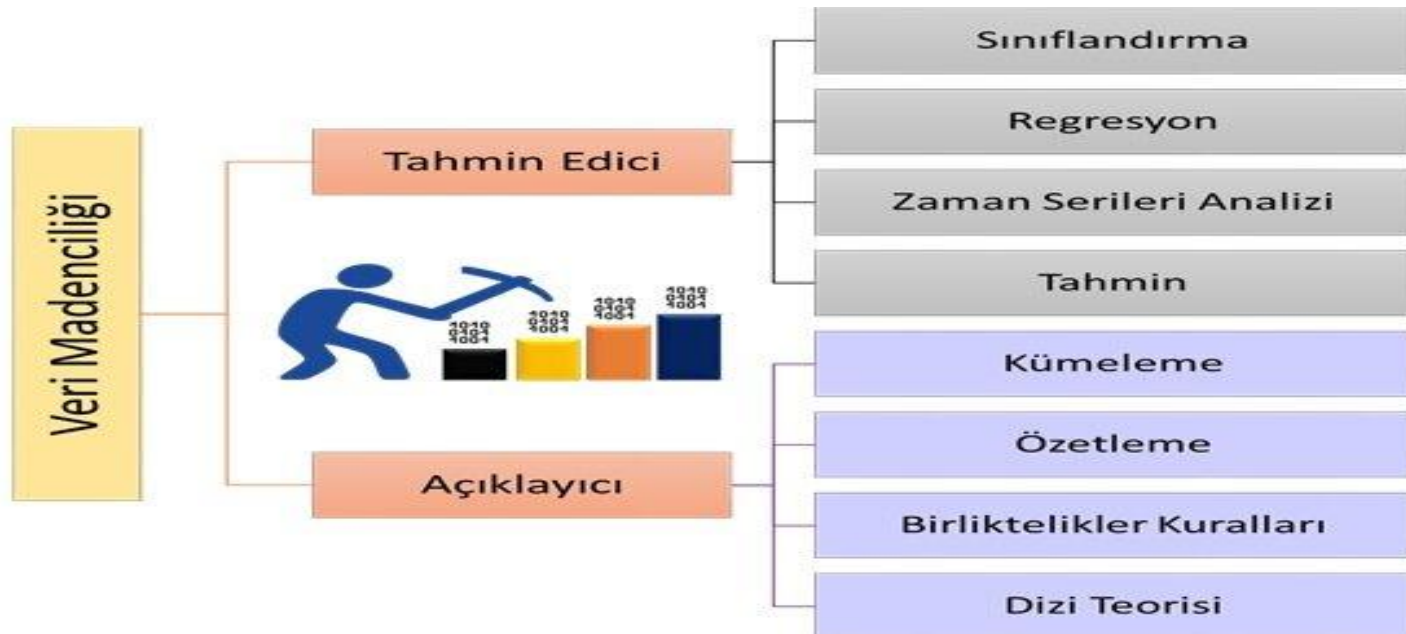
Veri Madenciliği Süreci

1. **Veri Toplama:** Ham verilerin toplanması (örneğin, sensörlerden, veritabanlarından).
2. **Veri Ön İşleme:** Veri temizleme, eksik verilerin tamamlanması, normalizasyon.
3. **Modelleme:** Algoritmaların uygulanması (sınıflandırma, kümeleme, regresyon).
4. **Değerlendirme:** Model performansının ölçülmesi ve iyileştirilmesi.
5. **Bilgi Keşfi:** Elde edilen bilgilerin raporlanması ve karar verme sürecine entegrasyonu.



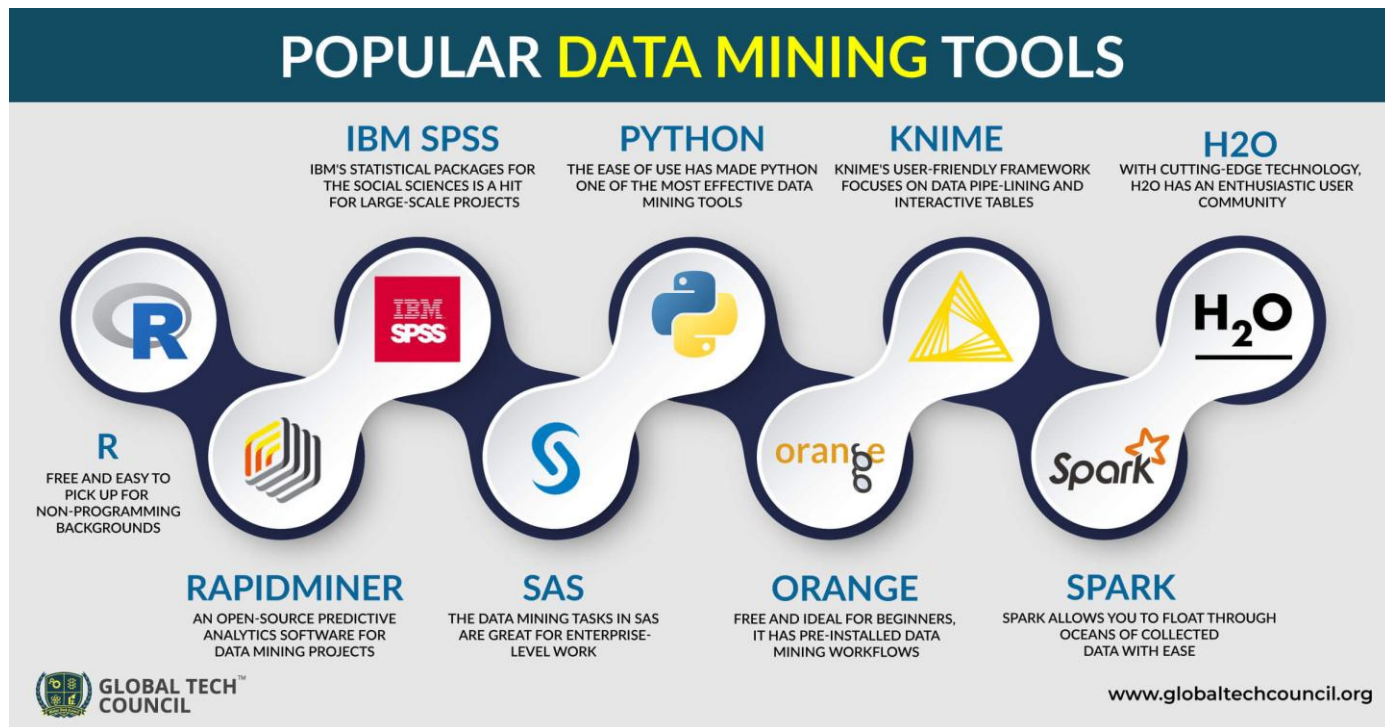
Veri Madenciliği Teknikleri

- **Sınıflandırma:** Veriyi önceden tanımlanmış sınıflara ayırma (örneğin, spam e-posta tespiti).
- **Kümeleme:** Benzer verileri gruptama (örneğin, müşteri segmentasyonu).
- **Regresyon:** Sürekli değerlerin tahmini (örneğin, enerji tüketimi tahmini).
- **İlişki Kurma:** Veri setindeki ilişkilerin keşfi (örneğin, market sepet analizi).



Veri Madenciliği Araçları

- **Python:** Scikit-learn, Pandas, NumPy, Matplotlib.
- **R:** İstatistiksel analiz ve veri madenciliği için kullanılan bir dil.
- **Weka:** GUI tabanlı veri madenciliği aracı.
- **Tableau:** Veri görselleştirme aracı.



Veri Madenciliği Uygulama Alanları

- ▶ Hangi promosyonu ne zaman uygulamalıyım?
- ▶ Hangi müşteri aldığı krediyi geri ödemeyebilir?
- ▶ Bir müşteriye ne kadar kredi verilebilir?
- ▶ Sahtekarlık olabilecek davranışlar hangileridir?
- ▶ Hangi müşteriler yakın zamanda kaybedilebilir?
- ▶ Hangi müşterilere promosyon yapmalıyım?
- ▶ Hangi yatırım araçlarına yatırım yapmalıyım?



Veri Madenciliği ile Makine Öğrenmesi Arasındaki Fark

Veri Madenciliği:

- **Tanım:** Büyük veri setlerinden anlamlı bilgiler çıkarma sürecidir. Veri madenciliği, veri setlerindeki örüntüleri, ilişkileri ve eğilimleri keşfetmeye odaklanır.
- **Amaç:** Daha önce bilinmeyen bilgileri keşfetmek ve bu bilgileri karar verme süreçlerinde kullanmak.
- **Teknikler:** Sınıflandırma, kümeleme, ilişki kurma, regresyon, veri özetleme gibi teknikler kullanılır.
- **Örnek:** Market sepet analizi (hangi ürünlerin birlikte satıldığını keşfetmek).

Makine Öğrenmesi:

- **Tanım:** Bilgisayarların verilerden öğrenmesini sağlayan bir yapay zeka alt alanıdır. Makine öğrenmesi, veriler üzerinde modeller oluşturarak tahminler yapmaya odaklanır.
- **Amaç:** Gelecekteki olayları tahmin etmek veya belirli bir görevi otomatikleştirmek.
- **Teknikler:** Denetimli öğrenme (sınıflandırma, regresyon), denetimsiz öğrenme (kümeleme), pekiştirmeli öğrenme gibi teknikler kullanılır.
- **Örnek:** Spam e-posta tespiti (e-postaları spam veya değil olarak sınıflandırmak).



Veri Madenciliği ile Makine Öğrenmesi Arasındaki Fark

Odak Noktası:

- Veri madenciliği, veri setlerindeki gizli örüntüleri keşfetmeye odaklanır.
- Makine öğrenmesi, verilerden öğrenerek tahminler yapmaya veya kararlar almaya odaklanır.

Kullanım Alanı:

- Veri madenciliği, daha çok keşif amaçlıdır (örneğin, müşteri segmentasyonu).
- Makine öğrenmesi, daha çok tahmin ve otomasyon amaçlıdır (örneğin, otomatik sınıflandırma).

Yöntemler:

- Veri madenciliği, istatistiksel yöntemler ve veri analizi tekniklerini kullanır.
- Makine öğrenmesi, algoritmalar ve modeller üzerine kuruludur.

Sonuç:

- Veri madenciliği, veri setlerindeki bilinmeyen bilgileri ortaya çıkarır.
 - Makine öğrenmesi, verilerden öğrenilen modellerle gelecekteki olayları tahmin eder.
-



Veri Madenciliği ile Makine Öğrenmesi Arasındaki Fark

- **Veri Madenciliği:** Veri setlerindeki **gizli örüntüleri keşfetmeye** odaklanır. Bu örüntüler, daha önce bilinmeyen ilişkiler, eğilimler veya anormallikler olabilir. Veri madenciliği, keşif amaçlıdır ve genellikle "neden" sorusuna cevap arar.
- **Makine Öğrenmesi:** Verilerden **öğrenilen modellerle tahminler yapmaya** odaklanır. Bu modeller, gelecekteki olayları tahmin etmek veya belirli bir görevi otomatikleştirmek için kullanılır. Makine öğrenmesi, "ne olacak" sorusuna cevap arar.



Örnek Uygulama: Enerji Tüketimi Tahmini

Makine Öğrenmesi

1. Problem Tanımı:

Bir fabrikanın veya bir binanın enerji tüketimini tahmin etmek istiyoruz. Bu tahmin, enerji maliyetlerini optimize etmek, enerji kaynaklarını daha verimli kullanmak ve olası enerji kesintilerini önceden öngörmek için kullanılabilir.

2. Veri Toplama:

- Enerji Tüketim Verileri:** Fabrikanın veya binanın geçmiş enerji tüketim verileri (örneğin, saatlik veya günlük enerji tüketimi).
- Hava Durumu Verileri:** Sıcaklık, nem, rüzgar hızı gibi hava durumu verileri (enerji tüketimi genellikle hava koşullarına bağlıdır).
- Üretim Verileri:** Fabrikada üretilen ürün miktarı veya çalışma saatleri (enerji tüketimi üretim miktarıyla doğru orantılı olabilir).
- Diğer Veriler:** Tatil günleri, özel etkinlikler gibi enerji tüketimini etkileyebilecek diğer faktörler.

3. Veri Ön İşleme:

- Veri Temizleme:** Eksik verilerin tamamlanması veya çıkarılması.
- Normalizasyon:** Verilerin aynı ölçeğe getirilmesi (örneğin, enerji tüketimi ve sıcaklık verilerinin farklı birimlerde olması).
- Özellik Mühendisliği:** Yeni özellikler oluşturma (örneğin, haftanın günü, mevsimsel etkiler).

4. Modelleme:

- Regresyon Modelleri:** Enerji tüketimini tahmin etmek için Lineer Regresyon, Ridge Regresyon veya Lasso Regresyon gibi modeller kullanılabilir.
- Zaman Serisi Modelleri:** ARIMA veya SARIMA gibi modellerle zaman içindeki enerji tüketimi eğilimleri analiz edilebilir.
- Makine Öğrenmesi Modelleri:** Destek Vektör Makineleri (SVM), Random Forest veya Gradient Boosting gibi modellerle daha karmaşık tahminler yapılabilir.

5. Değerlendirme:

- Model Performans Metrikleri:** Ortalama Kare Hatası (MSE), Ortalama Mutlak Hata (MAE) gibi metriklerle model performansı ölçülür.
- Çapraz Doğrulama:** Modelin farklı veri setleri üzerinde ne kadar iyi çalıştığını test etmek için kullanılır.

6. Sonuçlar ve Uygulama:

- Tahminler:** Model, gelecekteki enerji tüketimini tahmin eder.
- Optimizasyon:** Tahminler kullanılarak enerji maliyetleri optimize edilir veya enerji kesintileri önlenir.
- Raporlama:** Elde edilen sonuçlar, yönetim veya ilgili birimlerle paylaşılır.

Örnek Uygulama: Enerji Tüketimi Tahmini

Veri Madenciliği

1. Veri Keşfi ve Örüntü Bulma:

Amaç: Enerji tüketimi verilerindeki gizli örüntüleri keşfetmek.

Örnek Analizler:

- **Zamansal Örüntüler:** Enerji tüketiminin günün belirli saatlerinde veya haftanın belirli günlerinde artıp azaldığını keşfetmek.
- **Hava Durumu İlişkisi:** Sıcaklık, nem veya rüzgar hızı gibi hava durumu faktörlerinin enerji tüketimini nasıl etkilediğini analiz etmek.
- **Anormallik Tespiti:** Enerji tüketimindeki ani artış veya düşüşlerin nedenlerini keşfetmek (örneğin, ekipman arızaları veya özel etkinlikler).

2. İlişki Kurma ve Birliktelik Kuralları:

Amaç: Enerji tüketimi ile diğer faktörler arasındaki ilişkileri keşfetmek.

Örnek Analizler:

- **Üretim Miktarı ve Enerji Tüketimi:** Üretim miktarı arttıkça enerji tüketiminin nasıl değiştiğini analiz etmek.
- **Hava Durumu ve Enerji Tüketimi:** Sıcaklık artışı ile enerji tüketimi arasındaki ilişkiyi keşfetmek.

3. Kümeleme Analizi:

Amaç: Benzer enerji tüketimi davranışlarına sahip zaman dilimlerini veya bölgeleri gruplamak.

Örnek Analizler:

- **Zaman Dilimlerini Kümeleme:** Enerji tüketiminin benzer olduğu saatleri veya günleri gruplamak.
- **Bölgesel Kümeleme:** Farklı binaların veya fabrikaların enerji tüketim davranışlarını karşılaştırmak.

4. Veri Görselleştirme:

Amaç: Keşfedilen örüntüleri ve ilişkileri görsel olarak ifade etmek.

Örnek Görselleştirmeler:

- Enerji tüketiminin zaman içindeki değişimini gösteren çizgi grafikler.
- Hava durumu faktörleri ile enerji tüketimi arasındaki ilişkiyi gösteren scatter plotlar.
- Kümeleme sonuçlarını gösteren heatmap veya dendrogramlar.