



İstanbul
GEDİK
Üniversitesi



İstanbul
GEDİK
Üniversitesi

Veri Madenciliği

Öğr. Gör. Zeki ÇIPLAK

Veri Madenciliğinin İşletmelere Bazı Faydaları

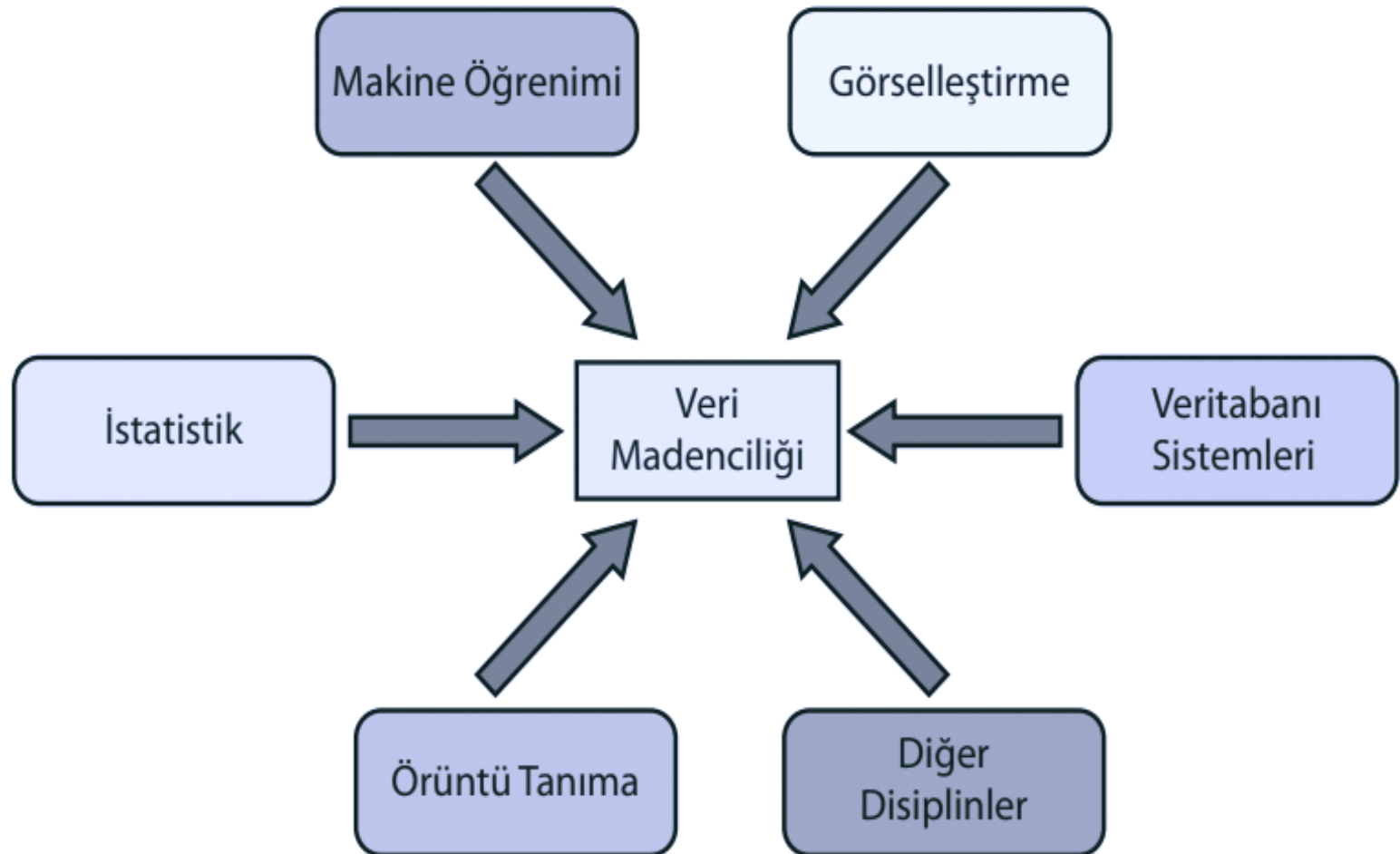
Veri Madenciliğinin İşletmelere Faydaları



- İşletmelerin **nerelerde zarar ettiklerini** daha iyi öğrenmelerini sağlar. Sahtekarlık/Dolandırıcılıkların ortaya çıkarılmasına yardımcı olur. (**Outlier Analizi**)
- İşletmelerin müşterilerini gruplara ayırmalarını ve her gruba ayrı özel kampanyalar yapabilmelerine imkan sağlar. (**Müşteri Segmentasyonu**)
- İşletmelerin çalışma şekillerini optimize eder ve daha fazla kâr etmelerini sağlar. Örneğin, hangi ürünün başka hangi ürünle birlikte satıldığı bilgisini verir. (**Birliktelik Analizi**)
- İşletmelerin geleceğe yönelik planlamalarına rehber olur.

Veri Madenciliğinin İlişkili Olduğu Disiplinler

Veri Madenciliğinin Etkileşimde Olduğu Disiplinler



Veri Madenciliğinin Etkileşimde Olduğu Disiplinler



- **Makine Öğrenimi:** Bilgisayarların da insanlar gibi öğrenebileceği düşüncesini temel alan bir disiplindir.
- **İstatistik:** Verilerin analizi ve değerlendirilmesi konusunda geçmişten günümüze yoğun olarak kullanılan bir disiplindir.
- **Görselleştirme:** Verilerin, tablolar ve grafikler gibi görseller yardımıyla sunulmasını ifade eder.
- **Veritabanı:** Verilerin yapılandırılmış halde depolandığı sistemlerdir.
- **Örüntü Tanıma:** Verilerin içinde, düzenli ve periyodik olarak tekrar eden ilişkileri bir model olarak kabul eden ve bu modelin (örüntünün) benzerlerini ya da en benzerini veritabanı içinden arama ve bulmaya yönelik teknolojidir.

Veri Madenciliği & Makine Öğrenmesi

Veri Madenciliği & Makine Öğrenmesi

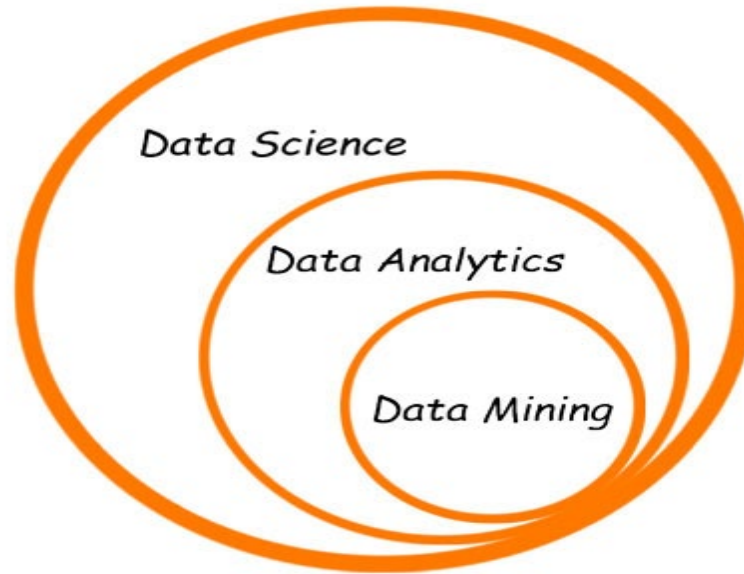


- Veri Madenciliği, Makine Öğrenmesi ile birbirine çok yakın iki alandır. Çoğu zaman aynı anlamda kullanılabilmektedir. Fakat aralarında bazı temel farklılıklar vardır.

Veri Madenciliği	Makine Öğrenmesi
Büyük miktarda veriden bilgi çıkarma işidir. Kararlara destek olur.	Algoritmalar kullanarak, verilerden bilgi üretmeyi amaçlar . Kararı kendi verir.
Odak noktası verilerdir .	Verilerden daha çok öğrenmeye odaklanır.
Uygulama noktasında, insan müdahalesine açıktır.	Uygulama noktasında, Otomatiktir . İnsan müdahalesine ihtiyaç duymaz.

Veri Madenciliği & Veri Bilimi

Veri Madenciliği & Veri Bilimi



- **Veri Bilimi**, büyük miktarda veri ile çalışmayı içeren bunları **öngörücü, kuralcı analitik modeller oluşturmak** için kullanan bir alandır.
- **Veri Madenciliği**, ise büyük bir veri kümesinden **önemli ve hayati bilgileri çıkarmak** için bir tekniktir.

Veri Madenciliği & SQL Sorguları

SQL Sorguları & Veri Madenciliği



- **Veritabanı Sorguları**

- Soyadı Öztürk olan müşterileri bulmak.
- Bir ayda 10000 liradan fazla harcama yapan kredi kartı sahiplerini bulmak.
- Önce kola, ardından cips satın alan müşterileri bulmak.

- **Veri Madenciliği Uygulaması**

- Kredi verilmesi az riskli olan kişileri bul. (Sınıflandırma)
- Harcama alışkanlığı benzer olan müşterileri bul. (Kümeleme)
- Ekmek ile birlikte sıkça satın alınan ürünleri bul. (Birliktelik Analizi)

Veri Madenciliğinin Uygulanma Alanları

Veri Madenciliği Nereelerde Uygulanır?



- Veri Madenciliğinin uygulanacağı alanlar, kesin çizgilerle ayrılamaz.
- Verinin doğru bir şekilde kaydedildiği ve veriler içerisinden faydalı bilgi çıkarılmak istenen her alanda, Veri Madenciliğini uygulayabilmek mümkündür.
- Veri Madenciliğinin en çok kullanıldığı alanlar; **pazarlama, finans, eğitim ve sağlık** alanlarıdır.

Pazarlama Alanındaki Uygulamalar

- Hangi müşterilere promosyon yapılacağının tespiti.
- Benzer özelliklere sahip müşteri gruplarının oluşturulması.
- Müşterilerin alışveriş sırasındaki örüntülerinin tespiti.
- Müşterilerin kaybedilmemesi için, nelerin yapılması gerektiğinin tespiti.
- Yeni müşterileri, hangi sebeplerin etkilediğini bulmak.

Finans Alanındaki Uygulamalar

- Kredi kartı dolandırıcılıklarının belirlenmesi.
- Kredi veya Kredi kartı verilebilecek kişilerin tespiti.
- Hisse senedi fiyatlarının tahmin edilmesi.
- Hangi araçlara yatırım yapılması gerektiğinin tespiti.

Eğitim Alanındaki Uygulamalar

- Eğitimdeki aksaklıkların tespiti.
- Öğrencilerin başarıları veya başarısızlıklarının sebeplerinin tespit edilmesi.
- Öğrenci başarısının nasıl arttırılabileceğinin tespiti.
- Özel veya farklı öğrencilerin tespiti.

Sağlık Alanındaki Uygulamalar

- Yeni ilaçların geliştirilmesi.
- Piyasada var olan ilaçların etkilerinin belirlenmesi.
- Hastalara uygulanan test sonuçlarının tahmin edilmesi.
- Hastalıkların önceden teşhis ve tedavi edilmesi.

Veri Madenciliğinin Sorunları

Veri Madenciliğinde Sorunlar

*Bu konuların dışında da karşılaşılabilecek birçok sorun olabilir.
Bunlar kesinlikle göz ardı edilmemesi gereken konulardır.*

- **Gizlilik ve Sosyal Haklar**

- *Kişilere ait verilerin toplanarak, kişilerden habersiz ve izinsiz olarak kullanılması.*
- *Veri madenciliği yöntemleri ile bulunan sonuçların izinsiz olarak açıklanması.*
- *Gizlilik ve veri madenciliği politikalarının düzenlenmesi.*

- **Bulunan sonuçlar bilinçsiz kullanılmamalı**

- *Veri Madenciliği geçmişteki olaylara bakarak örüntüler bulur ama gelecekteki olaylar farklıdır.*
- *İlişkiler her zaman nedenleri açıklamaz.*

Veri Madenciliğinde Veri Kaynakları

Veri Kaynakları



- Verilerin kaydedildiği dosyalar (csv, txt, doc, pdf vb.),
- Veritabanlarında depolanan veri kümeleri,
- Sensör verileri, Konumsal/Coğrafi veriler (Spatial data),
- Sosyal medya ağlarının verileri,
- Ses, video vb. gibi multimedya verileri,
- Devlet kurumları, dernekler vb. ait veriler...

Veri Kaynakları

- TÜİK: <https://data.tuik.gov.tr>
- İBB Veri Setleri: <https://data.ibb.gov.tr/dataset>
- Google Veri Setleri:
<https://datasetsearch.research.google.com>
- Kaggle Veri Setleri: <https://www.kaggle.com/datasets>