

A.Ü. ZİRAAT FAKÜLTESİ  
TARIM MAKİNALARI VE TEKNOLOJİLERİ  
MÜHENDİSLİĞİ BÖLÜMÜ  
TARIMDA VERİ BİLİMİNE GİRİŞ

Doç. Dr. Necati Çetin

1



# KAYNAKLAR

Python ile Makine Öğrenmesi  
(Prof. Dr. Engin Sorhun)

Makine Öğrenmesi Teorik Yönleri ve Python Uygulamaları  
(Dr. Öğr. Üyesi Sinan Uğuz)

Uygulamalarla Veri Bilimi  
(Doç. Dr. Deniz Kılınç – Nezahat Başeğmez)

Python ile Makine Öğrenmesi  
(Prof. Dr. Sadi Evren Şeker)

İnternet Kaynakları

necati.cetin@ankara.edu.tr  
necaticetin1990@gmail.com

# 1. GİRİŞ

- Veri, belirli bir yöntem ve daha fazla katma değer üreten bir biçimdir.
- Verinin bu seviyeye yükselmesi, büyük bir serüven ve şok edici bir metamorfozdur.

‘Veri yüzyılımızın petrolü...’

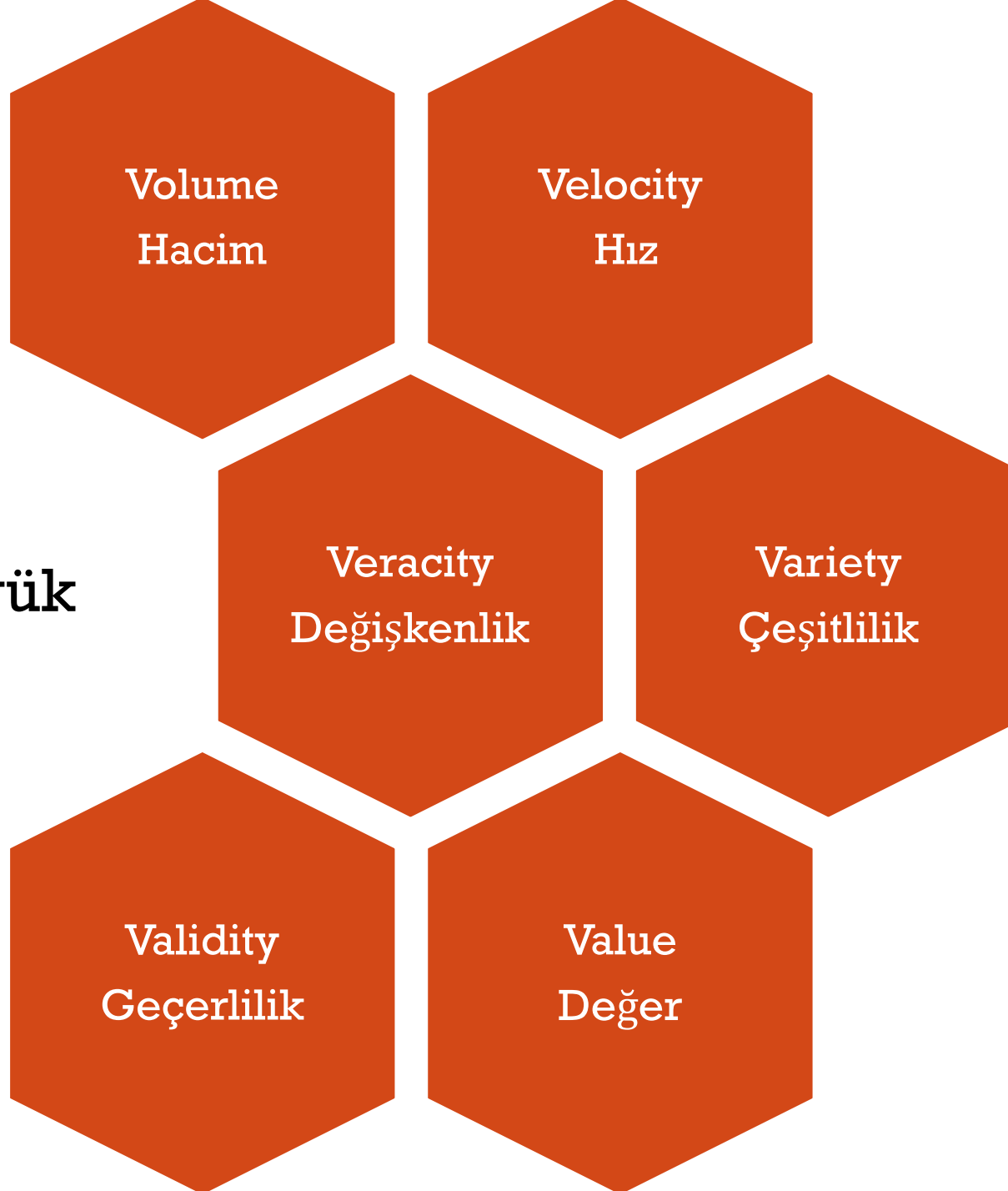


Farklı formatlarda her saniye güncellenen, adına 'Büyük Veri' dediğimiz evrendeki ham bilgiyi yönetmek, depolamak, işlemek, aktarmak, bilgiye dönüştürmek, karar alma süreçlerine dahil etmek, test etmek, tekrarlayarak geliştirip iyileştirmek...



- Verinin > Bilgiye, Bilginin > Kararlara dönüştüğü değer yaratma sürecinin aşamalarına '**Veri Bilimi**' adı verilir.

**Bir veri setini büyük yapan nedir ?**



**Cevap: 6V**





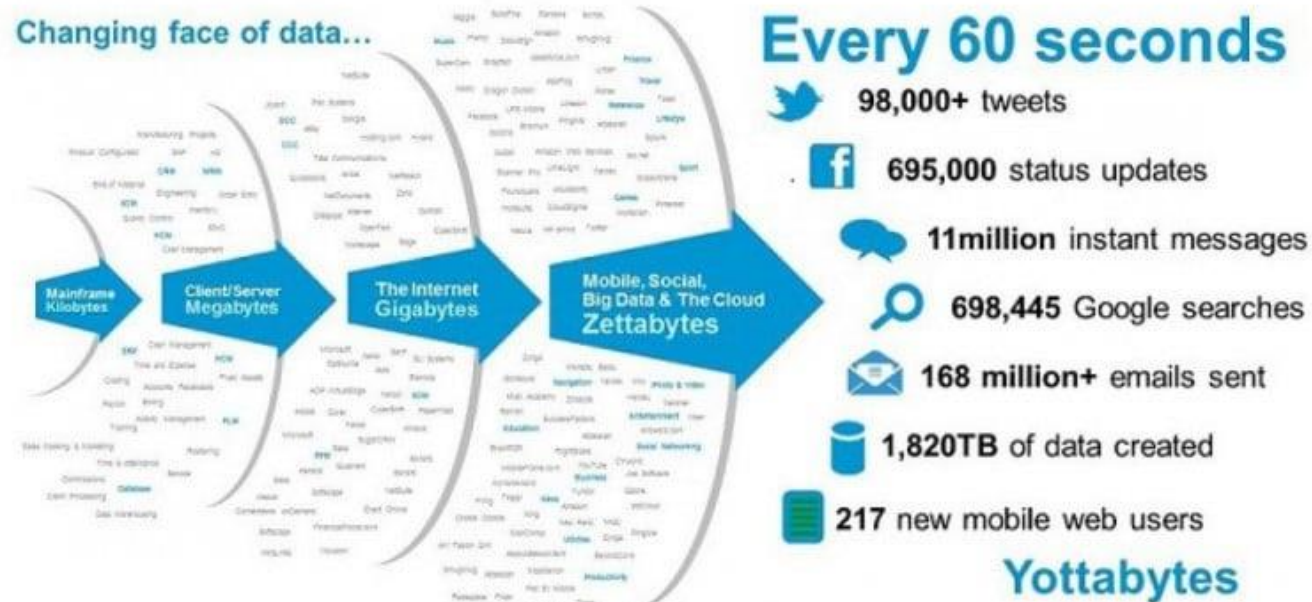
**Hacim**, her gün üretilen büyük miktarda veriyi ifade eder. Veri üretimi o kadar büyük ve karmaşıktır ki artık geleneksel veri işleme yöntemleri kullanılarak kaydedilemez veya analiz edilemez haldedir.

**Çeşitlilik**, veri türlerinin ve veri kaynaklarının çeşitliliğini ifade eder. Veriler her zaman yalnızca yapılandırılmış, yani sıra sütun düzeni içinde çoğunlukla hücrelerdeki sayılarla toplanmış şekilde değildir. Düzensiz veya yapılandırılmamış formatta ve türde bulunabilir.





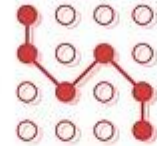
**Hız**, verilerin üretildiği, analiz edildiği ve yeniden işlendiği hızı ifade eder. Günümüzde veri akışı çoğunlukla eş zamanlı ya da gerçek zamanlı (real time) olarak bilinen bir saniyeden kısa bir süre içinde olup bitmektedir.



### Sources of Data Veracity



Statistical biases



Lack of data lineage



Software bugs



Noise



Abnormalities



Information Security



Untrustworthy data sources



Falsification



Uncertainty and ambiguity of data



Duplication of data



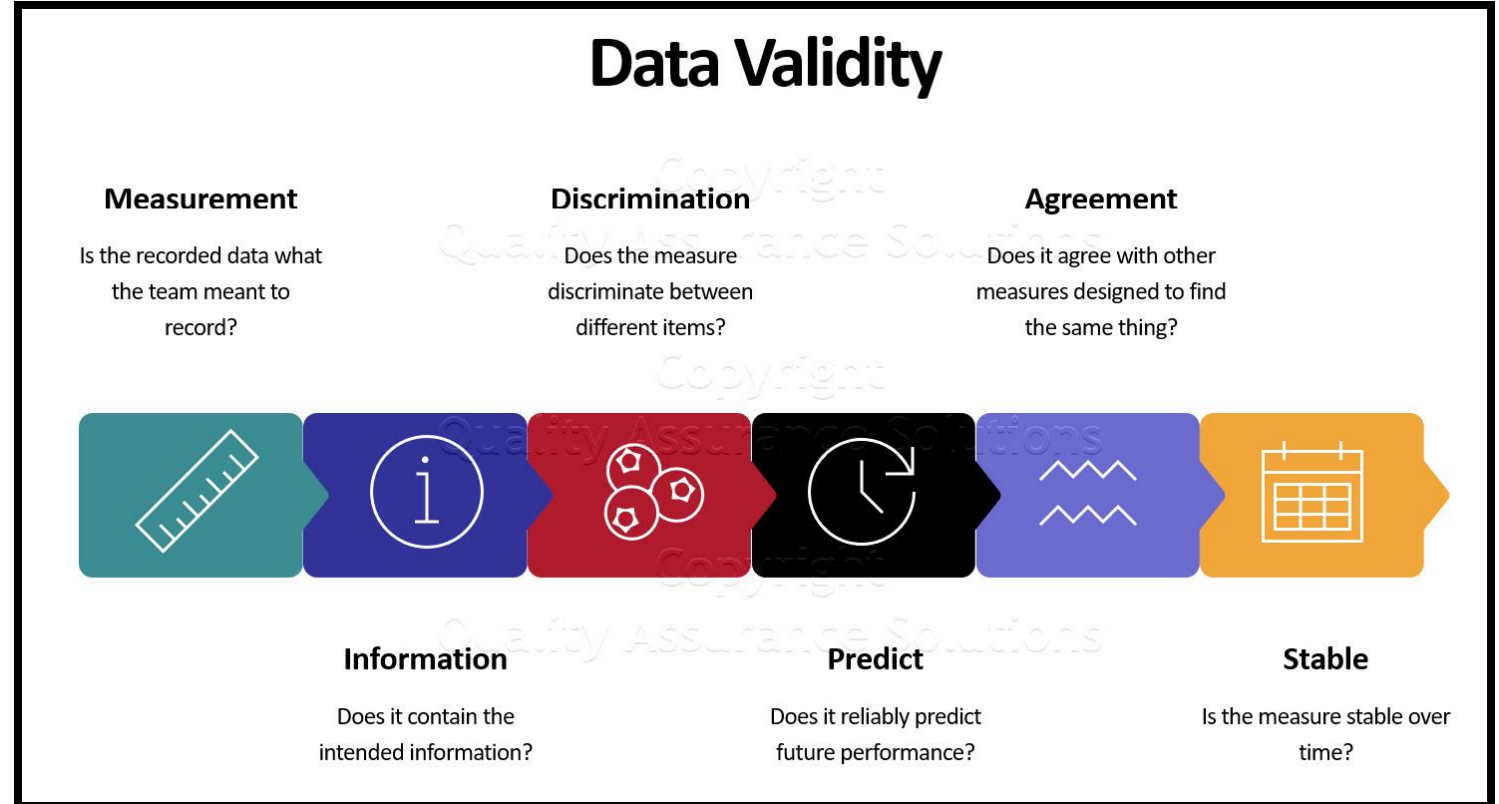
Out of date and obsolete data



Human error

**Değişkenlik**, veri setlerinin sürekli akışını, yeni verilerin eklenmesiyle güncellendiğini ifade eder.

**Geçerlilik**, verilerin yüksek hızlarda güncellenip sürekli daha yüksek hacimlere ulaşması sonucu ortaya çıkan kalite eksikliğini ifade eder. Güvenilir olmayan verilerle herhangi bir karar alınmaz. Karar alınsa da bu karar güvenilir olmaz.



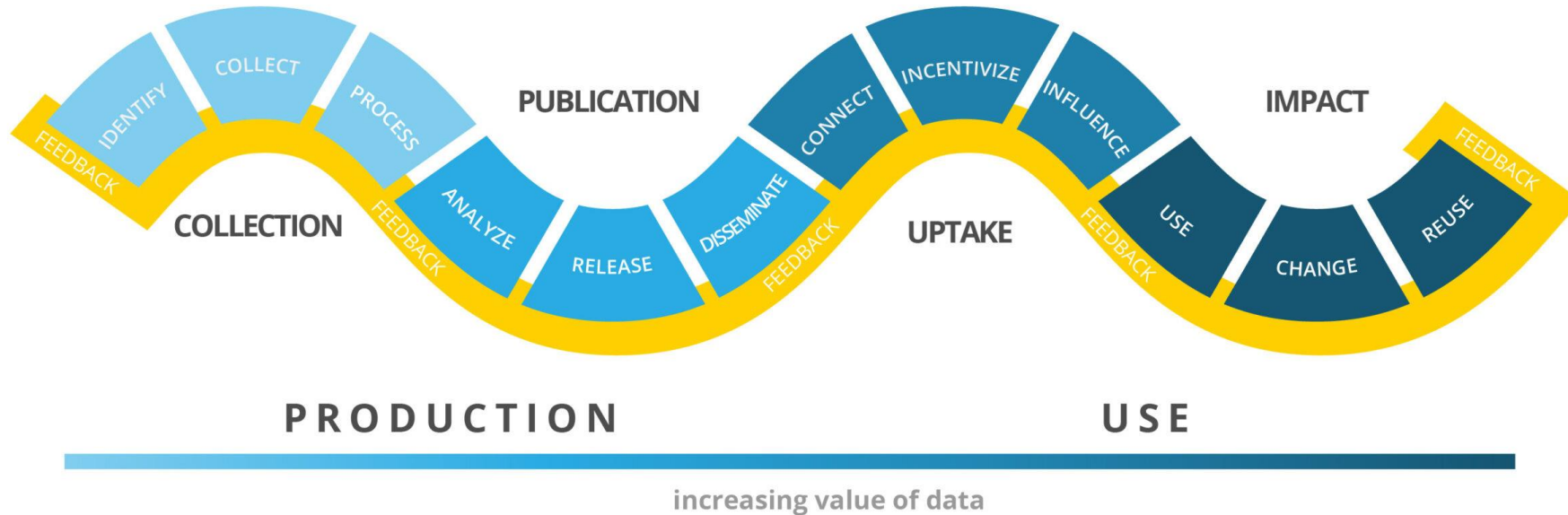


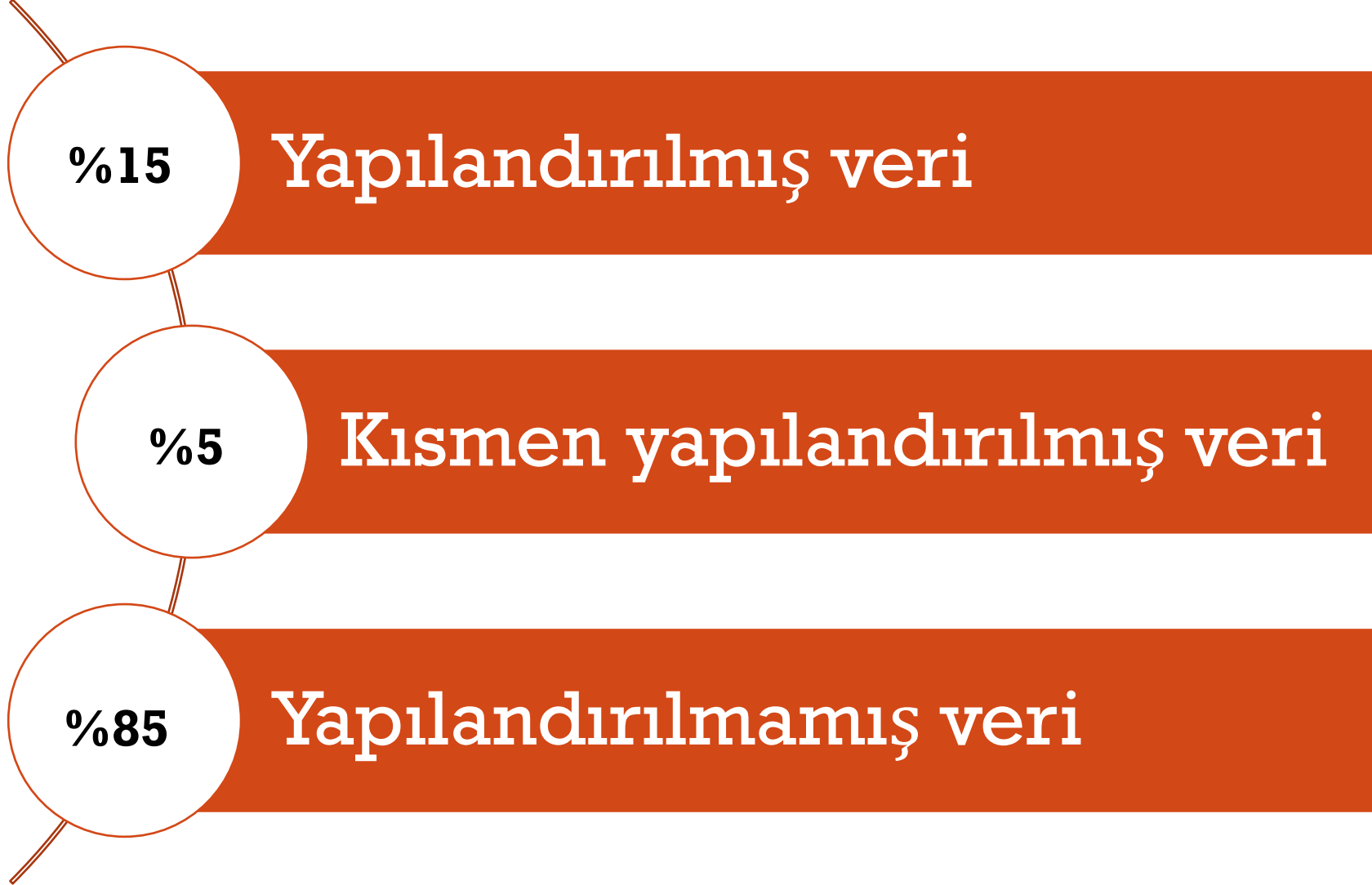
**Değer**, verinin onu kullananlar için katma değerini ifade eder. Karar alma süreçlerine katkısı, karar alıcıların etkin sonuç almaları, veriden elde edilen katma değere bağlıdır.

Birçok kurum ve kuruluş kendi veri platformlarına sahiptir.



## DATA VALUE CHAIN





*Bunlar aynı zamanda büyük veri türlerini ifade eder.*

**Yapılandırılmamış veriler**, sütun ve satırların bulunduğu tablolar halinde tutulan tüm verilerdir. (istatistik kurumlarının verileri)

**Kısmen yapılandırılmış veriler**, yapısal olduğu halde organize bir şekilde tutulmayan, düzensiz ya da farklı formatlarda tutulan verilerdir. (üniversite öğrencilerinin kayıt bilgileri XML formatındayken kütüphane bilgilerini Bibtex formatında olması)

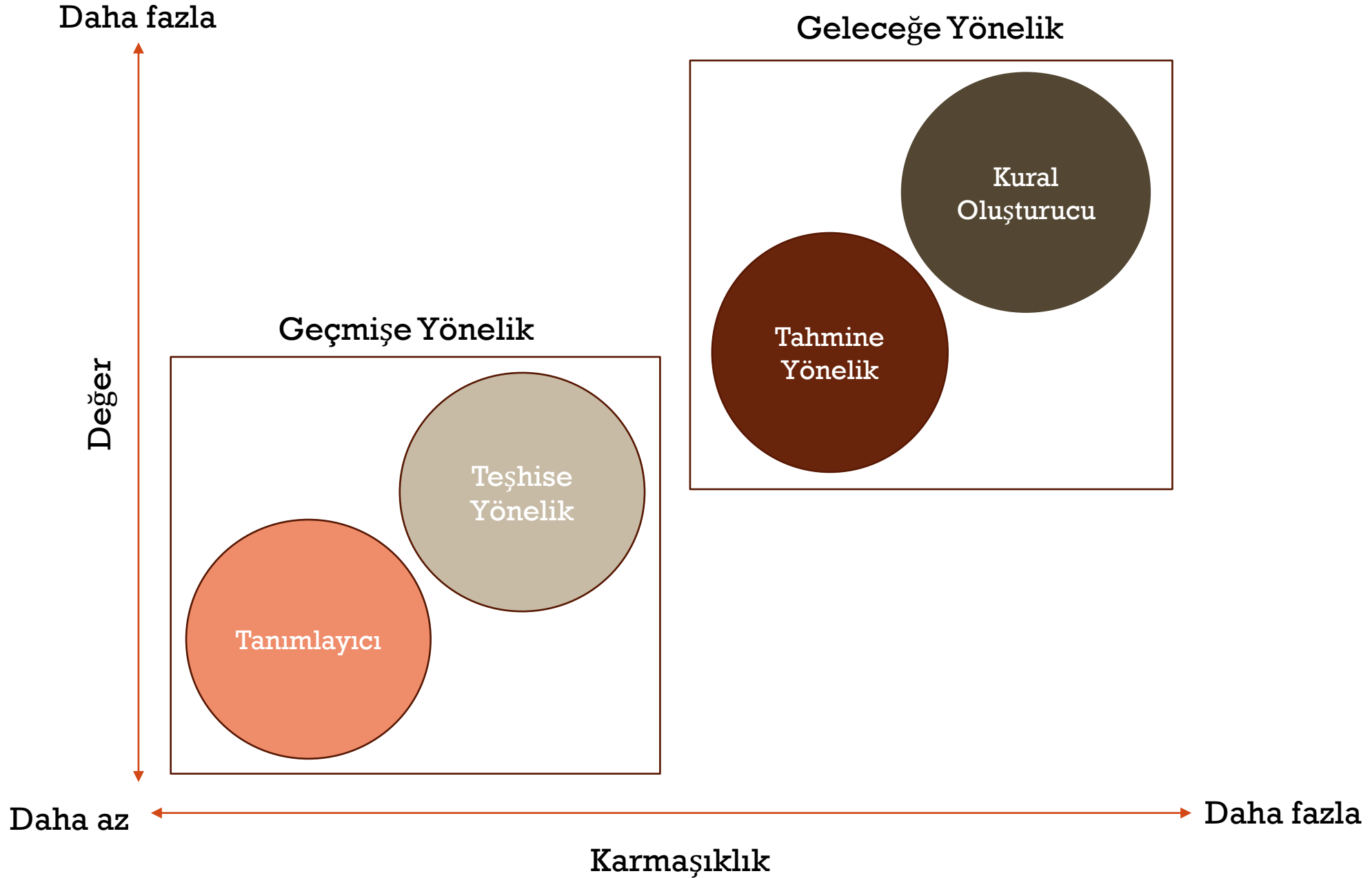


**Yapılandırılmış veriler** iki şekildedir. İlki makineler tarafından üretilenler; Uydu görüntüleri, fotoğraflar, videolar, ses kayıtları, radar, sonar, lidar kayıtları gibi.

İkincisi insanlar tarafından üretilenler; Metinler, raporlar, sosyal medya verileri, podcastler, mobil veriler (sms, konum bilgileri), web içerikleri gibi.

# Veriden değer nasıl üretilir ?

Bireyler ve kurumlar düzeyinde karar, eylem, politika ve iş yapış şeklinin gelişmesine yardımcı olacak, adına 'içgörü' (insight) denilen anlamlı bakış açıları ortaya çıkarmak için **veri toplama, işleme** ve **analiz** etme sürecine **analitik** denilmektedir. Veri analitiğinin dört türü bulunmaktadır. Bunların ikisi geçmişe, diğer ikisi geleceğe yöneliktir.





**Tanımlayıcı Analitik (Descriptive Analysis):** Bu geçmiş verileri, insanların kolayca okuyabileceği bir formda özetlemektedir. Veri toplama, veri madenciliği, kümeleme ve özet istatistikler (ortalama, standart sapma vb.) gibi tekniklerin tümü, geçmiş bir durumla ilgili analitik sağlamaya hizmet eder.

Bu bir şirketin geliri, satışları, hasat verileri, tohum boyutları, gübreleme miktarları, toprak özellikleri gibi raporların oluşturulmasına yardımcı olur.

**Teşhis analitiği (Diagnostic Analytics):** Tanımlayıcı analitik gibi teşhis analitiği de geçmişe odaklanır. Bu analizler bir şeyin neden gerçekleştiğini göstermek için neden ve sonuç arar. Amaç, nedenleri belirlemek için geçmiş olayları karşılaştırmaktır.

Aykırı değerlerin ortaya çıkarılması (outliner detection), regresyon analizi, keşfi veri analizi (exploratory data analysis), veri madenciliği, boyut azaltma vb. teknikler bu analizde sıkça kullanılır.

Belirli bir soruna derinlemesine bir bakış açısı sağlamak için teşhis analitiği kullanılır.

**Tahmine dayalı analitik (Predictive Analytics):** Bu tür analitik, geleceğe ilişkin potansiyel sonuçlar hakkında tahminlerde bulunmak için geçmiş ve mevcut verilere bakar.

Tahmine dayalı analitik, mevcut verileri analiz etmek ve gelecek hakkında tahminlerde bulunmak için veri madenciliği, yapay zekâ ve makine öğrenmesini kullanılır. Müşteri eğilimlerini, ürün verimini, yağış miktarını, pazar eğilimlerini vb. tahmin etmeye çalışır.



**Kural oluşturucu analitik (Prescriptive Analytics):** İstenen potansiyel (tahmin edilen) sonuçlara göre önerilen (öngörülen) eylemlerin belirlenmesine ve belirli bir sorunun çözümüne yönelik senaryo ve kural setleri, yol haritaları oluşturur. Hem tanımlayıcı, hem de tahmine dayalı analitik ile çalışır.

Çoğu zaman yapay zekâ, makine öğrenmesi ve derin öğrenme tekniklerine dayanır. Eylem ve olay ilişkilerini sürekli olarak analiz etmek ve en uygun çözümü önermek için veriler üzerinden sürekli kendini eğitir. Geri bildirim mekanizmaları aracılığıyla yol haritasında (iş stratejilerini) güncellemeler yapar.

# Veri Bize Ne Anlatır?

<b>Analitik</b>	<b>Soru</b>	<b>Örnek</b>
<b>Tanımlayıcı</b>	Ne oldu?	Geçen sene ne kadar ürün sattık?
<b>Teşhise Yönelik</b>	Neden oldu?	Geçen sene neden sadece A ürününü satabildik?
<b>Tahmine Yönelik</b>	Ne olacak?	Eğer A ürünü için reklam kampanyası düzenlersek ne kadar satarız?
<b>Kural Oluşturucu</b>	Nasıl Yapılmalı?	A ürününden 100.000 adet daha satmak için nasıl bir strateji uygulamalıyız?

## 2. YAPAY ZEKÂ NEDİR? BİLEŞENLERİ NELERDİR?

YAPAY ZEKÂ, 'Bir bilgisayarın veya bilgisayar kontrolündeki robotun çeşitli faaliyetleri zeki canlılara benzer şekilde yerine getirebilme yeteneğidir.'

Matematikçi Alan Turing İngiliz ordusu için geliştirdiği makine ile Nazilerin Enigma isimli bir makineyle ürettikleri şifreli haberleşmelerini kırıp savaşın seyrini değiştirmiştir. Bu isim aynı zamanda yapay zekânın öncüsü olarak kabul edilmektedir.

Turing test düzeneği hala günümüzde geçerliliğini sürdürmektedir.

Turing testinde ilginç olan bilgisayarların bir şey bilmeye ihtiyaçlarının olmamasıdır. Buna karşın makineler büyük miktarda veriyi, yüksek hızlarla işleyip, insan ile iletişim kurabilen ‘insanmış gibi’ davranabilen, insanı taklit eden simülasyon araçlarıdır.



# Yapay Zeka

Makine  
Öğrenmesi

Doğal Dil  
İşleme

Görüntü  
İşleme

Uzman  
Sistemler

Planlama ve  
Optimizasyon

Robotik

Konuşma  
Tanıma

Zekâ, insan kabiliyetleri açısından öğrenme, akıl/mantık yürütme, anlama/anlamlandırma, doğruyu yanlış ayırt etme, olaylar ve şeyler arasında ilgi/bağ kurma, somut ve soyut olanı ayırt etme gibi melekelerle ilişkilendirilmektedir.

Aynı zamanda teknik alanda veriden öğrenme, **hedef ve veri arasında mantıklı ilişkiler kurma**, gerekli ve değerli bilgiyi ayıklama, amaca uygun yol haritaları ortaya koyma, yeni durum ve bilgiye göre **önceki bilgiyi güncelleyip yol haritalarını geliştirme** gibi süreçlerle ilişkilendirilmektedir.

Yapay zekânın amacı insan zekasını kopyalamamak değildir. Bu alandaki çalışmaların amacı sadece *insanın öğrenme yöntemlerini taklit etmektir.*

Yapay zekâ doğru kurgulandığında, insanın günlük tekdüze işlerini daha hızlı ve verimli yapmasına sağlar. Böylece insana kendini gerçekleştirmek ve entelektüel hasletlerini geliştirmek için önemli bir fırsat sunar.

# MAKİNE ÖĞRENMESİ NEDİR? TÜRLERİ NELERDİR?

Makine öğrenmesi, **bilgisayarların veri setleri üzerinden öğrenimini mümkün kılan algoritmaların tasarım, geliştirme ve uygulama** süreçlerini konu edinen yapay zekânın bir alt dalıdır.

Makine öğrenmesi veriden en iyi sonuçları elde etmek için güçlü bir teknik çerçeve sunmaktadır.

Makine öğrenmesi, bir bilgisayarın açık komutlar kullanmadan belirli bir görevi gerçekleştirebildiği süreçtir. Bu durumu, geçmiş verilerde bulunan istatistikî modellere dayanarak elde etmektedir.

# Makine Öğrenmesi Algoritmalarının Türleri





Makine öğrenmesinde iki tür veri vardır:

- 1) Etiketli veriler (labeled)
- 2) Etiketlenmemiş veriler (unlabeled)

## Etiketli veriler (labeled)

Hem girdi hem de çıktı parametreleri belli bir düzen içerisinde makine (bilgisayar) tarafından okunabilen verilerdir.

Ancak, başlangıçta verileri etiketlemek için çok fazla emek gerekir.

## Etiketlenmemiş veriler (unlabeled)

Girdi ve çıktı parametrelerinden birisinin ya da her ikisinin de makine/bilgisayar tarafından okunamadığı verilerdir.

Bu, insan emeğine olan ihtiyacı ortadan kaldırır, ancak karmaşık çözümler gerektirir.

# DENETİMLİ ÖĞRENME

Bu yöntemde, makine öğrenmesi algoritması etiketli veriler üzerinde eğitilir.

Bu yöntemin işe yaraması için verilerin doğru bir şekilde etiketlenmesi gerekmesine rağmen, denetimli öğrenme doğru ayarlandığında çok güçlü sonuçlar vermektedir.

Makine öğrenmesinde girdi parametrelerine **öznitelik** ya da **öznitelik değişkeni**, çıktı parametrelerine ise **hedef değişkeni** adı verilir.

Öznitelikler, *girdi görevi gören* bağımsız değişkenlerdir.

*Hedef (çıktı)* ise bağımlı değişkendir.

Veri setindeki her bir gözlemin öznitelik değerlerine karşılık gelen hedef değişken değeri ile etiketlenir.

Hedef değişken aslında gözlemlerin etiketidir.

Öznitelik sayısı arttıkça veri setinin boyutu da artar.



Çıktı

Girdi

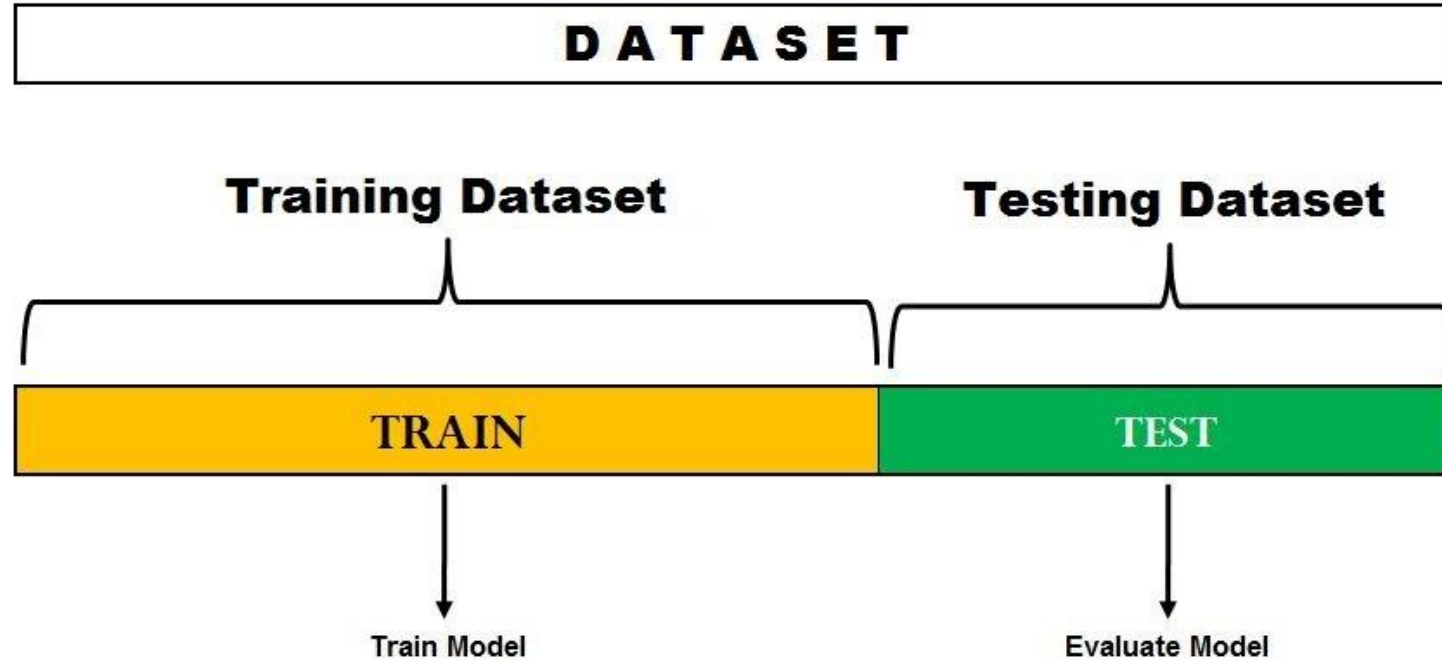
Gözlem	Hedef	Öznitelik 1	Öznitelik 2	Öznitelik 3
1	*	*	*	*
2	*	*	*	*
3	*	*	*	*
.				
.				
.				
N	*	*	*	*

Gözlem Sayısı

Denetimli öğrenmede etiketli veri setinin bir kısmı algoritmanın eğitilmesi için ayrılır. Buna EĞİTİM SETİ denir.

Algoritmanın eğitim seti üzerinden öğrenmesi gerçekleştirilir.

Veri setinin kalan kısmına da TEST SETİ denir.



Test seti karşılaşılabacak yeni durumları temsil eder.

Eğitilen algoritmanın yeni durumları tanımlayıp tanımlayamadığı test seti ile değerlendirilir.

Veri setinin bu şekilde ikiye ayrılmasına eğitim-test seti bölümlenmesi (train-test split) denir.

%80'e %20,

%70'e %30,

%66'ya %34 en yaygın bölümlenmelerdir.

Denetimli öğrenmeyi hedef değişkenin türüne bağlı olarak (sürekli değerler, sınıf veya kategori) genel olarak ikiye ayırabiliriz.

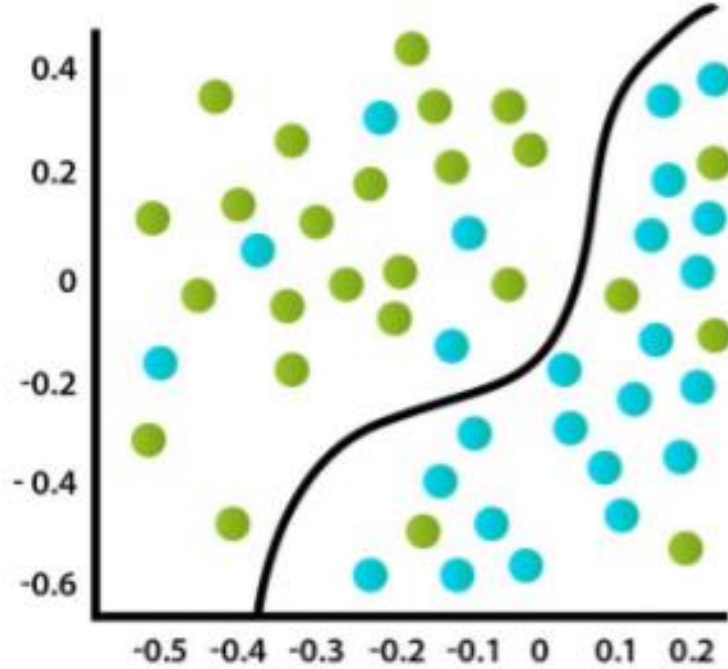
## Sınıflandırma ve Regresyon

**Sınıflandırmada**, öz niteliklere karşılık gelen hedef değişken sınıf ya da kategorik değerlerle etiketlenmiştir. (var ya da yok, erkek veya kadın, doğru veya yanlış, a çeşidi veya b çeşidi vb.)

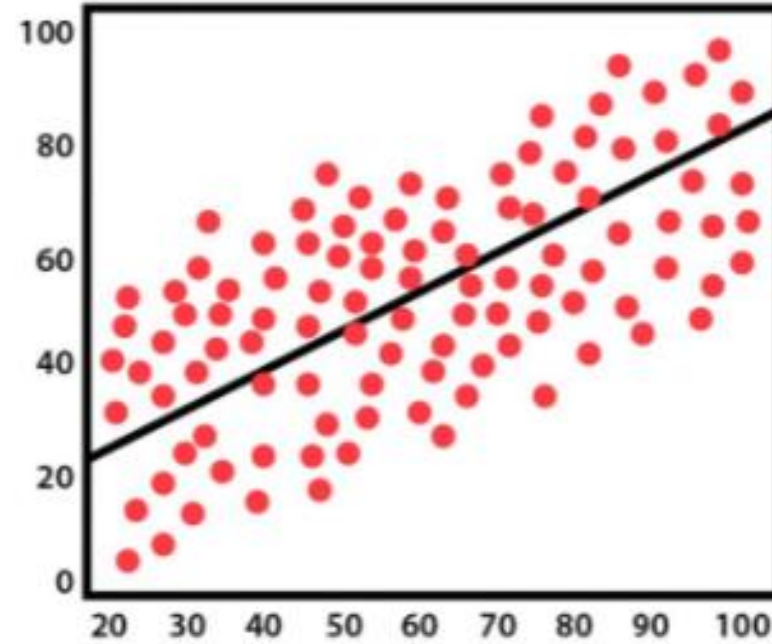
**Regresyonda** ise, öz niteliklere karşılık gelen hedef değişken sürekli değerlere sahiptir. (fiyat, maaş, yaş, verim, kütle vb.)

**Sınıflandırmanın amacı** gözlemlerin ya da veri noktalarının sınıflarını ya da kategorilerini tahmin etmektir.

**Regresyonun amacı** gözlemlerin ya da veri noktalarının etiket değerlerini tahmin etmektedir.



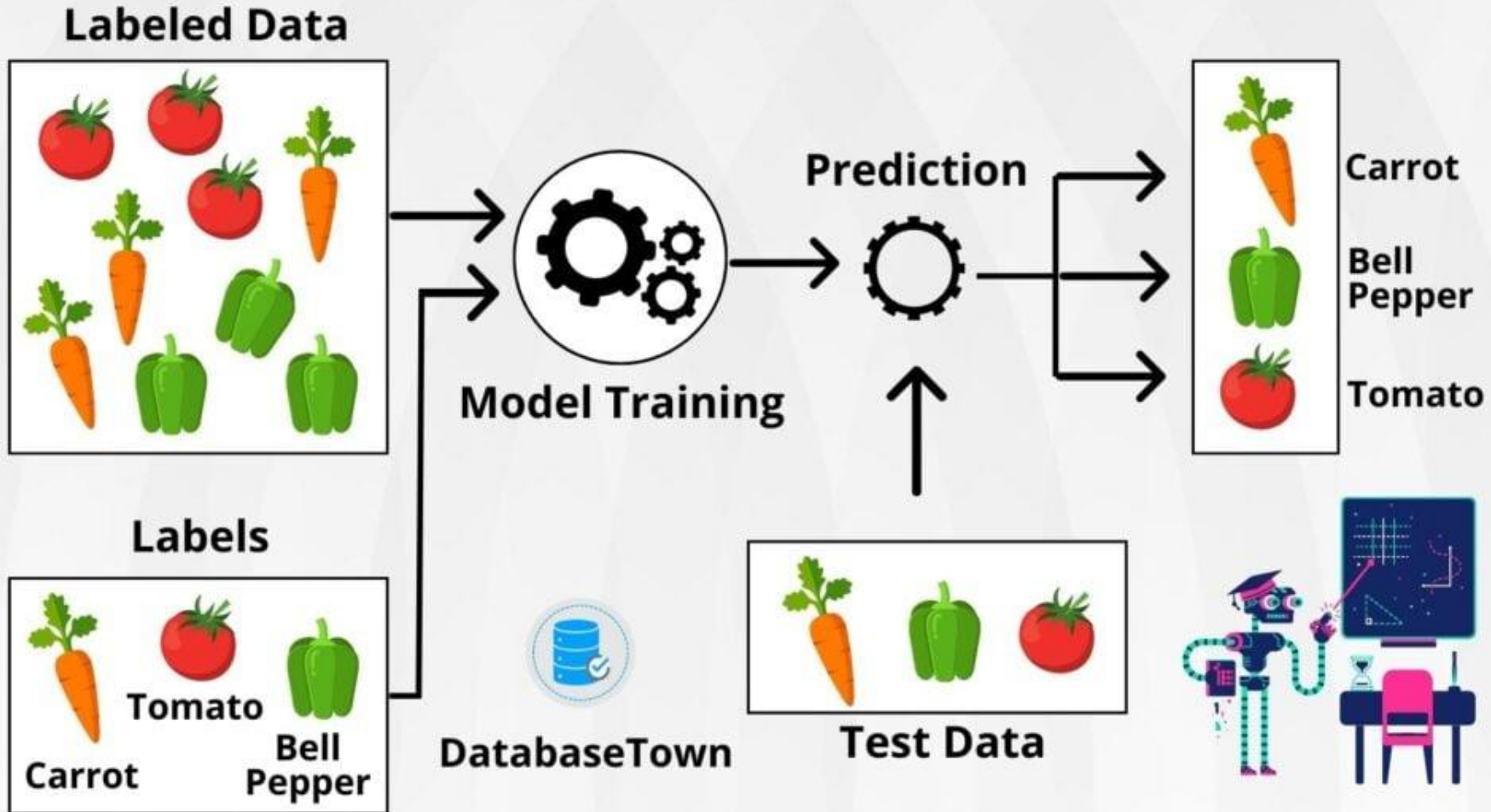
Sınıflandırma



Regresyon

# SUPERVISED LEARNING

Supervised machine learning is a branch of artificial intelligence that focuses on training models to make predictions or decisions based on labeled training data.





# DENETİMSİZ ÖĞRENME

Denetimli öğrenmede etiketler, algoritmanın veri noktaları arasındaki ilişkisinin yapısını bulmasına izin verir.

Ancak, denetimsiz öğrenmenin üzerinde çalışacak etiketleri yoktur.

Bu da gizli yapıların oluşmasına neden olur.

Veri noktaları arasındaki ilişkiler, insanların hiçbir girdi gerektirmeden, algoritma tarafından soyut bir şekilde algılanır.

Denetimsiz öğrenme, denetimli makine öğrenmesini aksine, *etiketlenmemiş verilerden* bilgi elde etme avantajına sahiptir.

Bu, veri setinin makine tarafından okunabilir hale getirmek için insan emeğinin gerekli olmadığı ve algoritma tarafından çok daha büyük veri setleri üzerinden öğrenme yapılabileceği anlamına gelir.

Çıktı

Girdi

Gözlem	Hedef	Öznitelik 1	Öznitelik 2	Öznitelik 3
1	*	*	*	*
2	*	*	*	*
3	*	*	*	*
.				
.				
.				
N	*	*	*	*

Gözlem Sayısı

Denetimsiz makine öğrenmesi veri noktaları arasında gizli kalmış yapıları, analitik bağları, ilişkileri ortaya çıkarmayı ve bunları bir fonksiyonla tanımlamayı amaçlar.

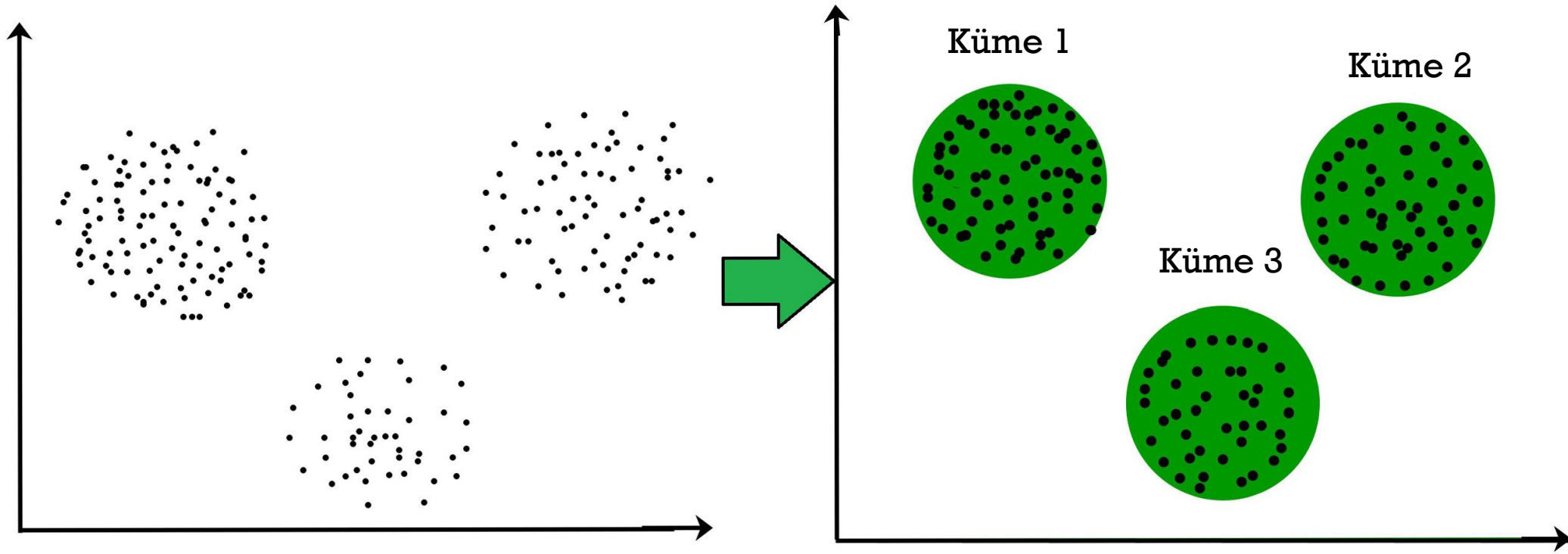
- 1) Kümeleme,
- 2) Birliktelik Kuralları,
- 3) Boyut Azaltma gibi alt türleri bulunmaktadır.

## Kümeleme

Bir veri setinde benzer özellikler gösteren verilerin gruplara ayrılmasına denir.

Aynı küme içinde benzerlikler fazla, kümeler arası benzerlikler azdır.

Kümeleme algoritmaları **müşteri segmentasyonu, pazar segmentasyonu, görüntü işleme ve makine görüşü** gibi alanlarda yaygın olarak uygulanır.





## Birliktelik Kuralları

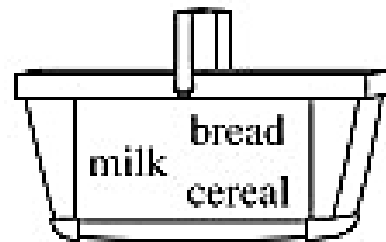
Büyük veri setleri içinde yer alan öge kümeleri arasındaki ilginç ilişkileri ve ilişki yapılarını bulmaya yarayan bir denetimsiz makine öğrenmesi yöntemidir.

En yaygın uygulama örneği Pazar Sepeti Analizidir.

**Pazar Sepeti Analizi**, perakendecilerin, insanların sık sık birlikte satın aldıkları ürünler arasındaki ilişkileri belirlemelerine olanak sağlar.

Which items are frequently purchased together by customers?

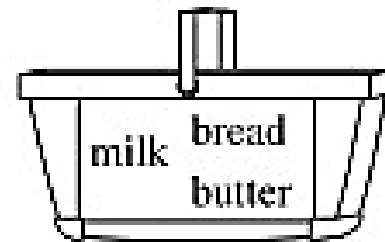
### Shopping Baskets



Customer 1



Customer 2



Customer 3



Customer  $n$

Market Analyst

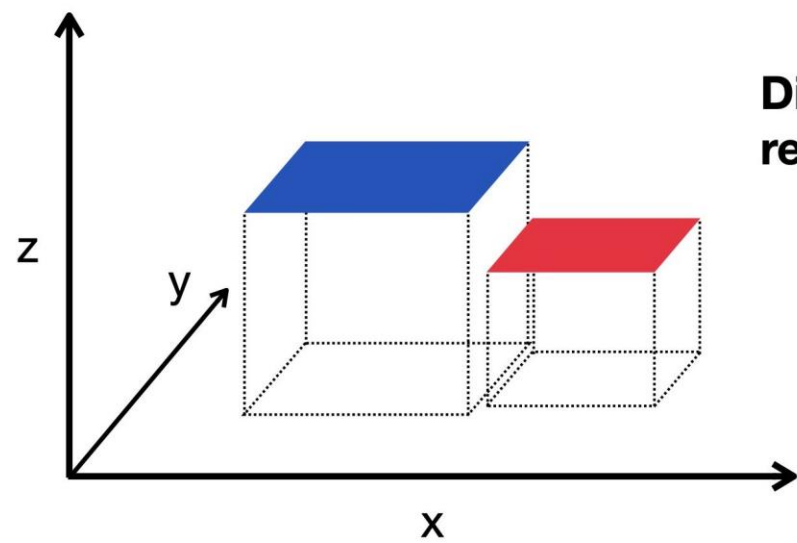
## Boyut Azaltma

Çoklu verilerde hedef değişkenin bir yönünü işaret eden, tanımlayan özniteliklerin belirlenmesidir.

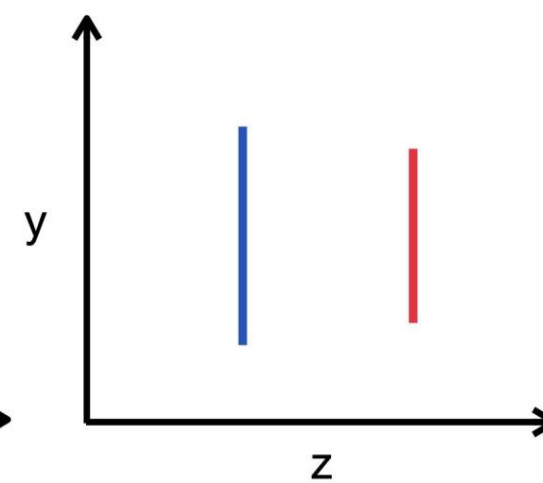
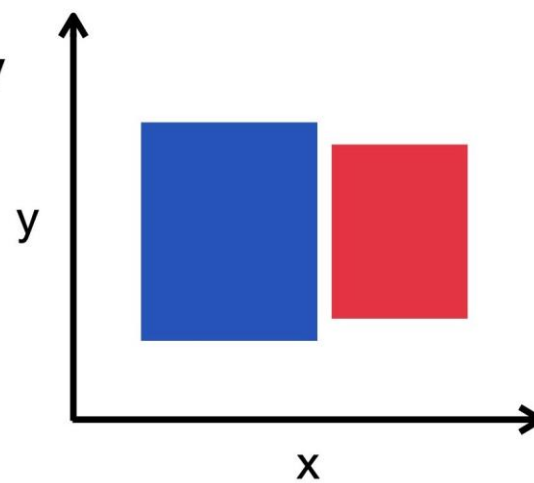
Yani her bir öznitelik verisi tahmin modeline bir boyut katmaktadır.

Bu avantaj sağlamaktadır. Çünkü, bir olaya ne kadar çok farklı açıdan bakarsak onu o kadar iyi değerlendirebiliriz.

Algoritma bir olayı ne kadar farklı bir boyuttan inceler ve öğrenirse o kadar iyi tanır ve önüne yeni bir durum (gözlem) geldiğinde onu nereye konumlandıracağını daha iyi öngörebilir.



**Dimensionality  
reduction**



# PEKİŞTİRMELİ ÖĞRENME

Doğrudan insanların hayatlarındaki verilerden nasıl öğrendiklerinden ilham alır.

Makine, deneme yanılma yoluyla sürekli olarak kendini eğittiği bir ortama (çevreye) maruz kalır.

Karşılaşılan her yeni durumda belli bir davranış kalıbı ya teşvik edilir ya da cezalandırılır.

Teşvik edilen davranış, elverişli çıktılar ile pekiştirilmiş olur.

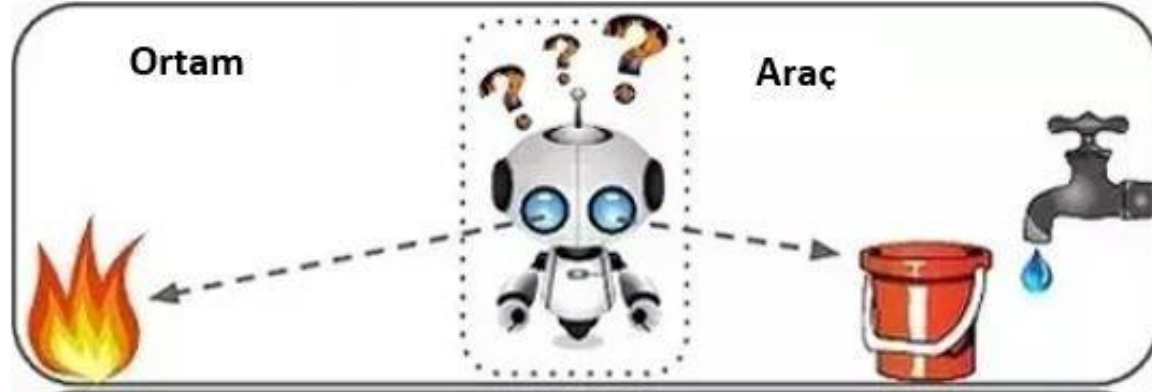
Bilgisayar oyunlarında pekiştirmeli öğrenme algoritmaları kullanılır.  
Bazı oyunların ödül-ceza kurgusu, çocuklara ve gençlere yanlış davranışlar kazandırmasından ve bağımlılık yapmasından dolayı eleştiri konusu olmaktadır.

Bu algoritmayı kullanarak, model belirli kararlar vermek üzere eğitilir.

*Yöntem şu şekilde çalışır: makine, deneme yanılma yoluyla sürekli olarak kendini eğittiği bir ortama (çevreye) maruz kalır.*

Bu makine geçmiş deneyimlerinden öğrenir ve doğru iş kararları vermek için mümkün olan en iyi bilgiyi yakalamaya çalışır.





1

Gözlemler

2

Hareketi seç



3

Hareket

4

Ödüllendirir ya da cezalandırır



5

Planı güncelle

6

Bulduğumuz en uygun politikaya kadar tekrarla