

Makine Öğrenmesi



Gazi Üniversitesi YZT

Neler Konuşacağız?

- Makine Öğrenmesi Nedir?
- Makine Öğrenmesi Uygulamaları
- Gözetimli ve Gözetimsiz Öğrenme
- Metodoloji (CRISP-DM)
- Veri Tipleri (Kategorik - Sayısal)
- Kategorik Veri Tiplerinin Sayısal Veri Tipine Dönüştürülmesi
- Kukla Değişken Tuzakı (Dummy Variable Trap)
- Basit Lineer Regresyon
- Korelasyon Matrisi
- Çoklu Lineer Regresyon

Makine Öğrenmesi Nedir?

Makine Öğrenimi (ML), tükettikleri verilere göre öğrenen ya da performansı iyileştiren sistemler oluşturmaya odaklanan bir yapay zeka (AI) alt kümesidir.

Yapay zeka, insan zekasını taklit eden sistemler veya makineler anlamına gelen kapsamlı bir terimdir.

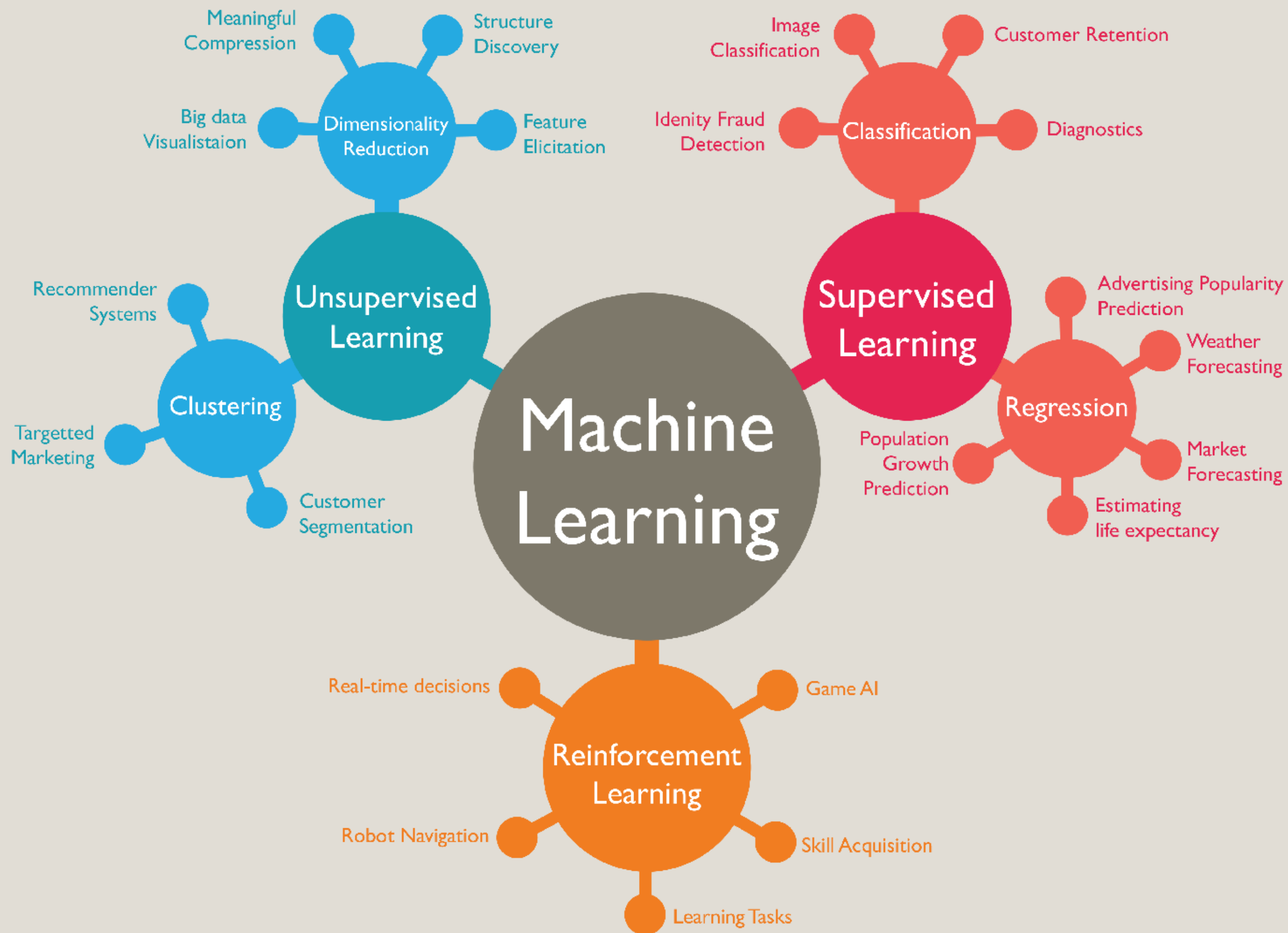
```
54 padding: 4px 6px;  
55 text-align: left;  
56  
57 &:hover {  
58     color: $c-link-hover;  
59 }  
60  
61 &.selected {  
62     background-color: $c-action  
63     color: white;  
64 }  
65  
66 .amount {  
67     float: right;  
68     font-weight: bold;  
69 }  
70  
71 &.last-child {
```

People who
use libraries



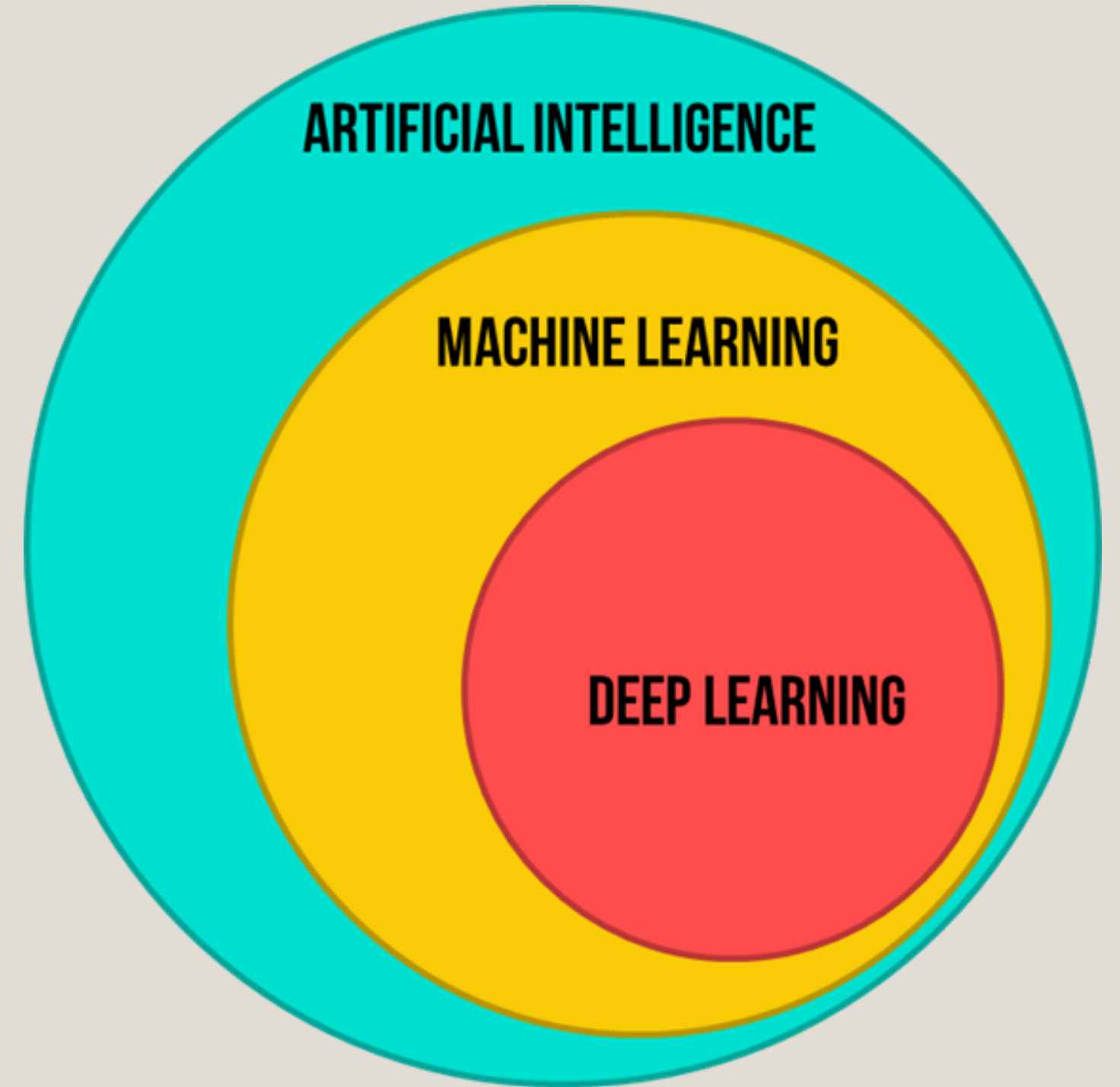
People who
make
libraries





ML vs. AI vs. DL

**Makine Öğrenmesi
bir yapay zeka alt
kümesi iken Deep
Learning bir Makine
Öğrenmesi alt
kümesidir.**



Makine Öğrenmesi Uygulamaları

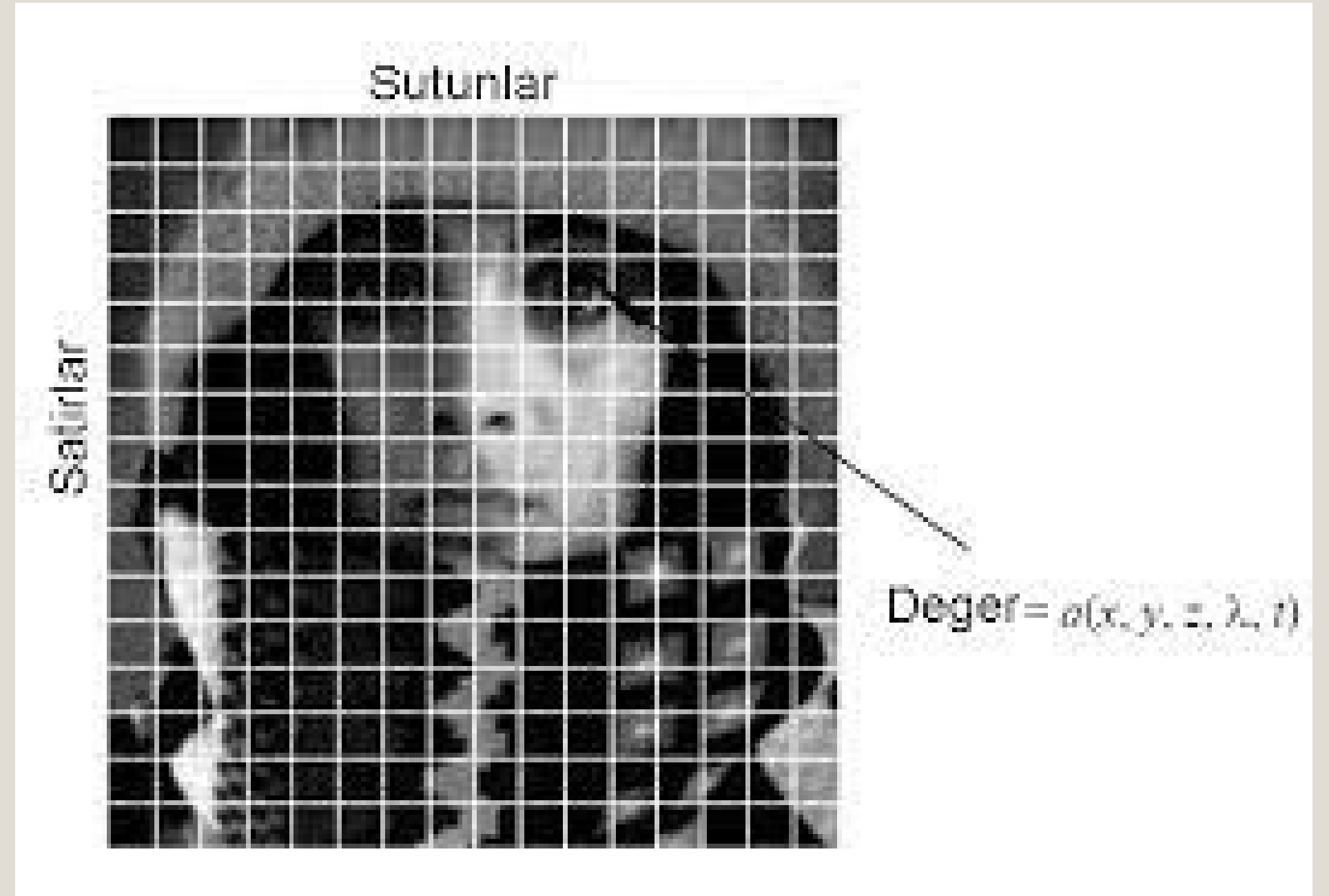
- Bilgisayar ile görü ve nesne tanıma
(Computer vision and object recognition)

Bilgisayar üzerinden yapılan yapay zekalı yazılımın belli hareketleri veya olaylara karşı bir aksiyon oluşturması veya alan oluşturmasını sağlamaktır.



● Görüntü İşleme (Image processing)

Görüntü işleme, bilgisayarların görüntüler ve videolardaki nesneleri ve kişileri tanımlamasını ve anlamasını sağlamaya odaklanan bir bilgisayar bilimi alanıdır.



● Sanal Gerçeklik (Virtual Reality)

**Sanal gerçeklik,
görüntüleri iki boyutlu
ekranlardan algılamak
yerine üç boyutlu
gerçeklikte
deneyimlemeye yarayan
bir teknolojidir.**



● Artırılmış Gerçeklik (Augmented Reality)

Artırılmış gerçeklik, gerçek dünyadaki çevrenin ve içindekilerin bilgisayar tarafından üretilen; çeşitli veriler ile zenginleştirilerek meydana getirilen canlı veya dolaylı fiziksel görünümüdür.



- **Pazarlama ve reklam**
- **Müşteri ilişkileri yönetimi**
- **Sağlık sektörü**
- **Sahtekarlık yakalama**
- **Arama motorları**
- **Endüstri 4.0**
- **IoT (Internet of Things)**

Gözetimli ve Gözetimsiz Öğrenme (Supervised and Unsupervised Learning)

● Gözetimli Öğrenme (Supervised Learning)

Makineyi bağımlı ve bağımsız değişkenin bir arada bulunduğu, etiketlenmiş veriler kullanarak eğitmeye verilen isimdir.

● Gözetimsiz Öğrenme (Unsupervised Learning)

Gözetimsiz öğrenme, etiketlenmemiş veri kümelerini analiz etmek ve kümelemek için yapay öğrenme algoritmalarını kullanır.

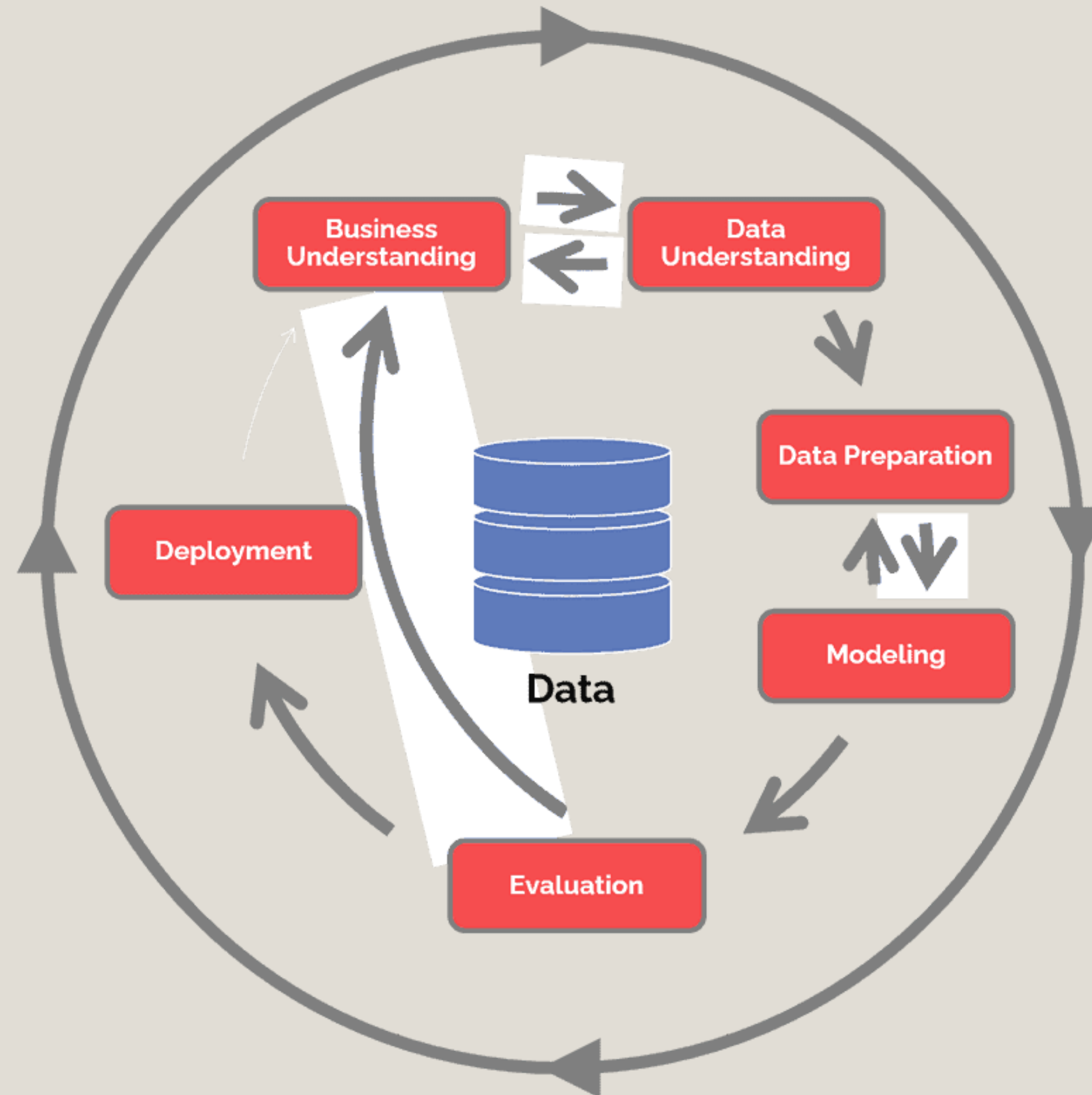
Sınıflandırma

- Verilerin etiketi vardır.
- Verileri bir gruba dahil etmek için bir kural oluşturulmasını bekler.
- Veri setini eğitim ve test olarak ayırmak gereklidir.

Kümeleme

- Verilerin etiketi yoktur.
- Verileri bir gruba, yakınlığa,benzerliğe,alakaya, hiyerarşiye göre dahil eder.
- Verideki örüntülerini ve yapıları tespit eder.

Metodoloji (CRISP - DM)



Veri tipleri



Kategorik veriler

- **Ordinal veri tipi** : Kendi aralarında sıralanabilen fakat ölçülemeyen veri tipleridir.
- **Nominal veri tipi** : Hem kendi aralarında sıralanamayan hem de ölçülemeyen veri tipleridir.

Sayısal veriler

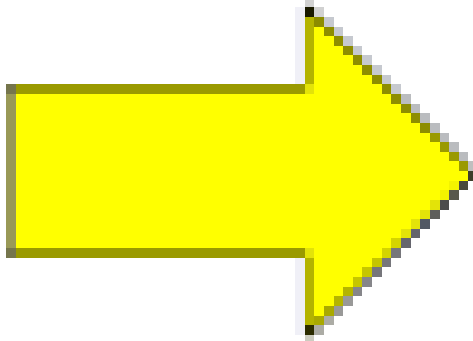
- **Oransal veri tipi** : 4 işleme imkan tanıyan verilerdir. "Ratio" olarak da bilinir.
- **Aralık belirten veri tipi** : Bu veri tipleri toplama çıkarma gibi işlemlere imkan tanırken çarpma bölme gibi işlemlere imkan tanımaz.

Kategorik Verileri Sayısal Veriye Dönüştürmek

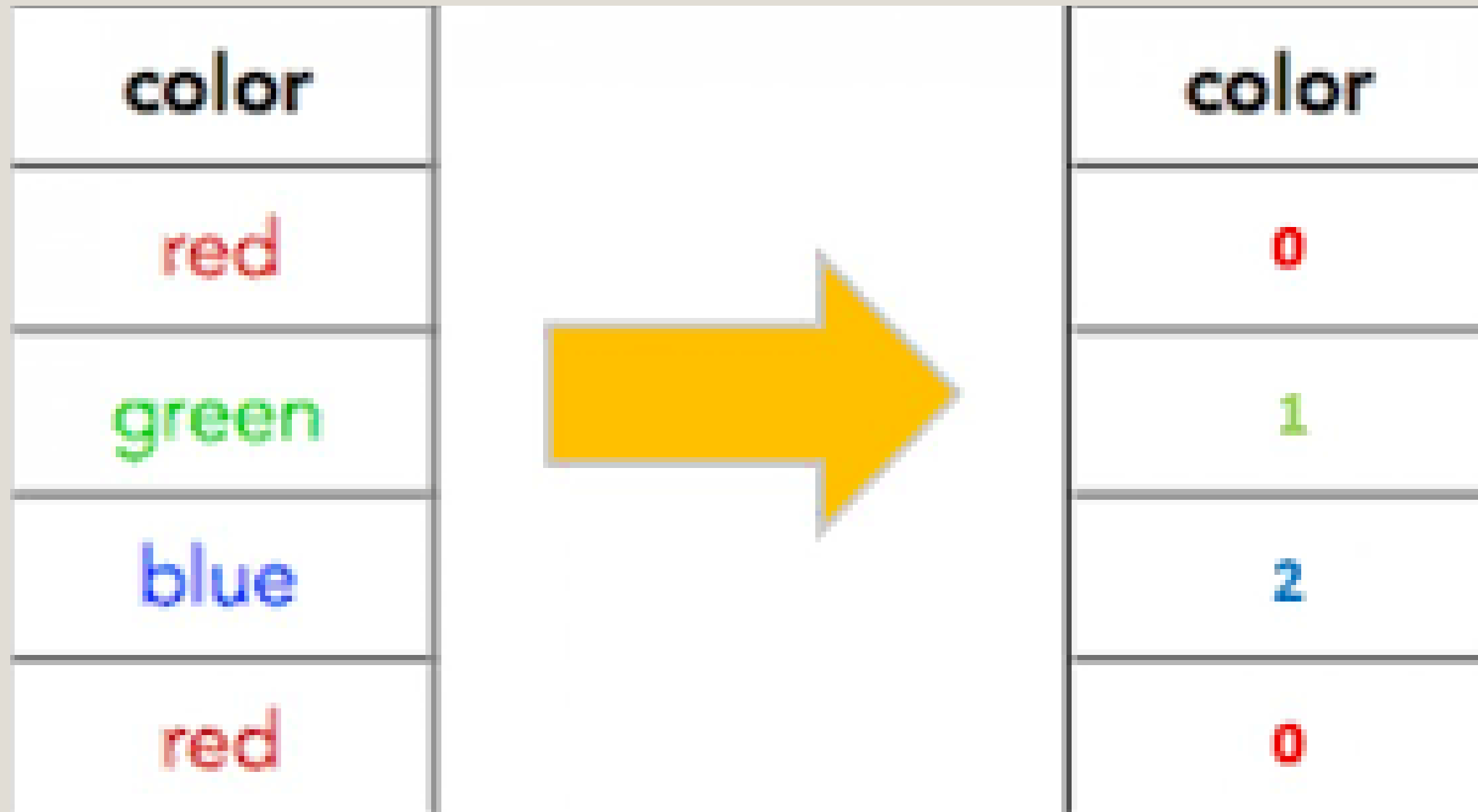
- Kategorik veriler karşılaştırma yapmaya uygun olmadığından makine öğreniminde kullanılmaz. Bu yüzden veriyi uygun olan sayısal veri tipine dönüştürmemiz gerekir.
- Bunun için Label Encoder, Ordinal Encoder ve One Hot Encoder gibi modüller kullanılır.

ulke	bov	kilo	vas	cinsivet
tr	130	30	10	e
tr	125	36	11	e
tr	135	34	10	k
tr	133	30	9	k
sp	129	38	12	e
sp	180	90	30	e
sp	190	80	25	e
us	175	90	35	e
us	177	60	22	k

One Hot Encoder

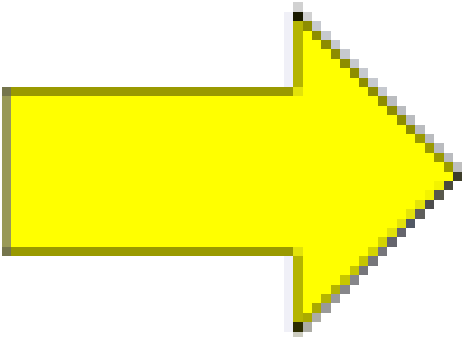
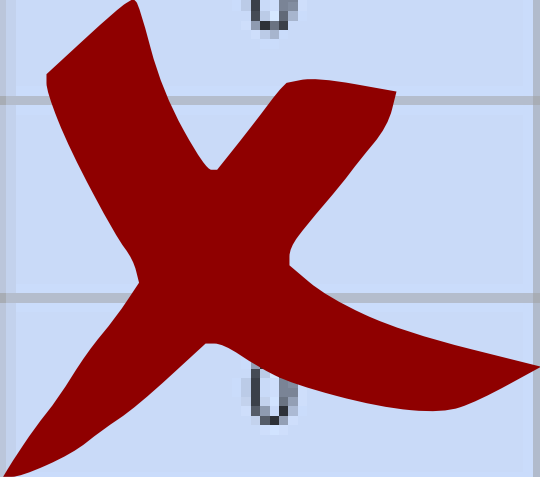
Color		Red	Yellow	Green
Red		1	0	0
Red		1	0	0
Yellow		0	1	0
Green		0	0	1
Yellow		0	0	1

Label Encoder



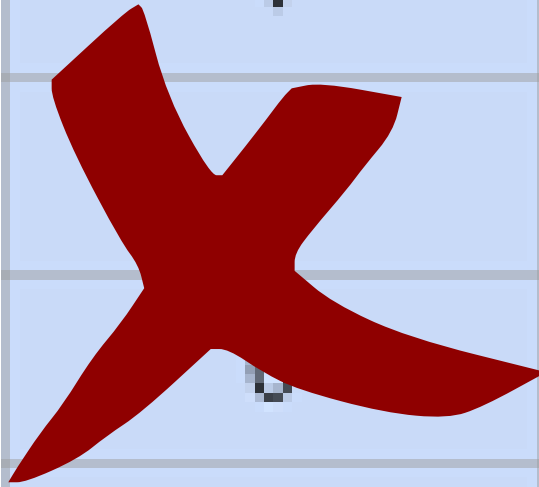
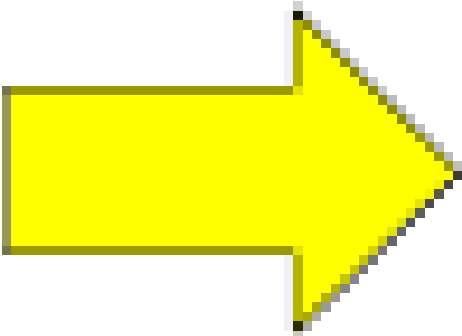
Kukla Degisken ve Kukla Degisken Tuzagi

(Dummy Variable Trap)

Color		Red	Yellow	Green
Red				
Red		1	0	0
Yellow		1	0	
Green		0	1	
Yellow		0	0	
				1

Kukla Degisken ve Kukla Degisken Tuzagi (Dummy Variable Trap)

Color		Red	Yellow	Green
Red				
Red		1	0	0
Yellow			0	0
Green			1	0
Yellow		0	0	1

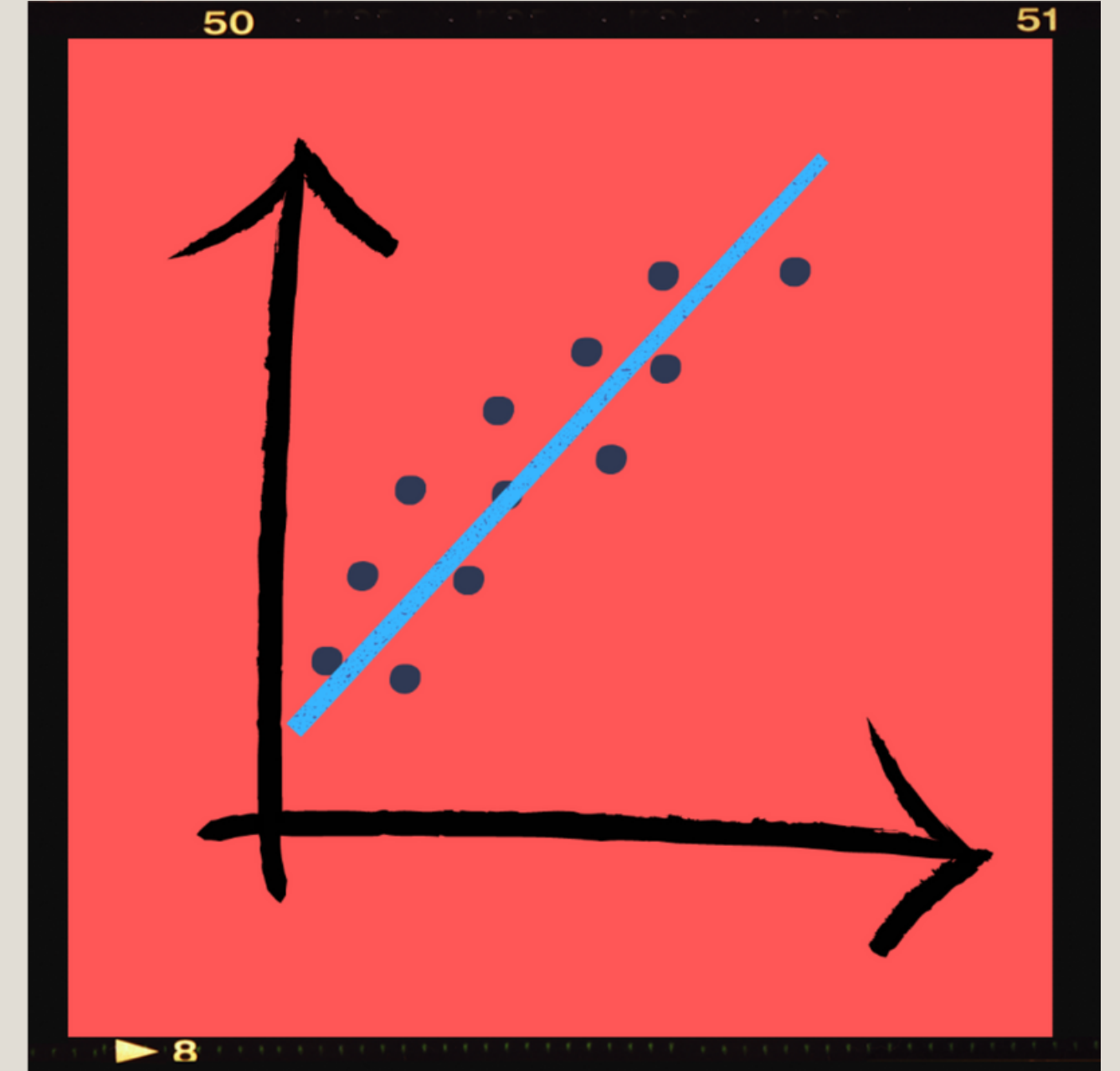


The diagram illustrates the Dummy Variable Trap. It shows a mapping from a categorical variable 'Color' to a set of dummy variables (Red, Yellow, Green). The first table lists the categories: Red, Red, Yellow, Green, and Yellow. The second table shows the corresponding dummy variable values. However, the second table contains a large red 'X' over the first three rows, indicating that the coding is incorrect or leads to a trap. The values in the second table are: Red (1, 0, 0), Yellow (0, 0, 0), Green (0, 1, 0), and Yellow (0, 0, 1).

Kukla Degisken ve Kukla Degisken Tuzagi (Dummy Variable Trap)

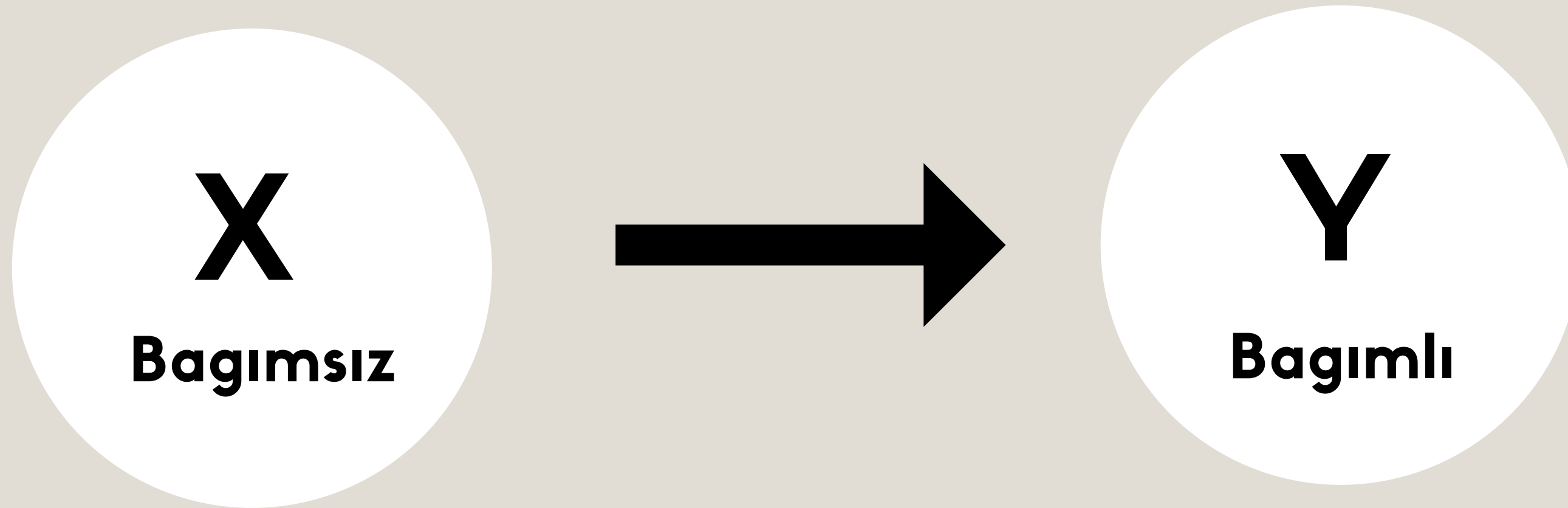
Color		Red	Yellow	Green
Red				
Red		1	0	0
Yellow		1		0
Green		0	1	0
Yellow		0	0	1

Tahmin Algoritmalarına Giriş



Lineer Regresyon

Basit Lineer Regresyon



1 bağımlı ve 1 bağımsız değişken bulunan değişkenler arasındaki ilişkinin doğrusal olduğu regresyondur.

Belirttigimiz üzere regresyon,değişkenlerden birisi (x) bilindiginde diger deęişkenin (y) alabilecegi deęerin tahmin edilmesidir.

$$\hat{y} = a + bx$$

The diagram illustrates the components of the regression equation $\hat{y} = a + bx$. Red arrows point from each term to its corresponding label below it:

- \hat{y} points to **Bagımlı Degisken** (Dependent Variable).
- a points to **Regresyon Sabiti** (Regression Constant).
- b points to **Regresyon Katsayısı** (Regression Coefficient).
- x points to **Bagımsız Degisken** (Independent Variable).

$$\hat{y} = a + bx$$

**Bagımlı
Degisken**

\hat{y} ,her bir x
degerine göre
tahmin edilen ve
bagımlı(sonuç)değiş
ken olan y
degerini ifade
etmektedir.

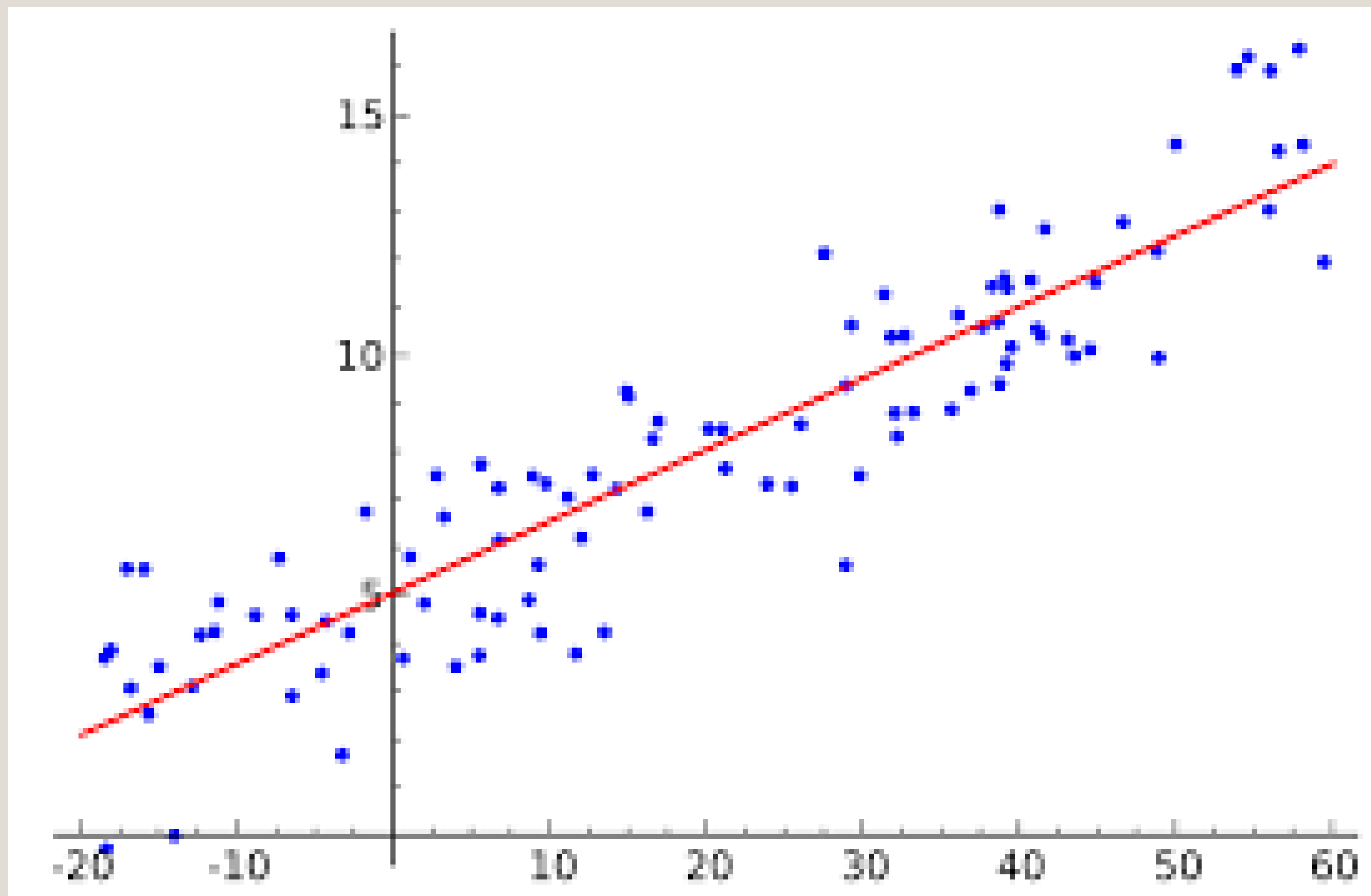
**Regresyon
Sabitı**

a ise yine tahmini y
degerinin bulunması
için x'e eklenmesi
gerekten sabit degerdir.

**Regresyon
Katsayısı**

b,tahmini y degerinin
bulunması için x'in
çarpılması gereken
katsayı degeridir.

**Bagımsız
Degisken**



Python ile Basit Lineer Regresyon Kodu

Çoklu Doğrusal Regresyon (Multiple Linear Regression)

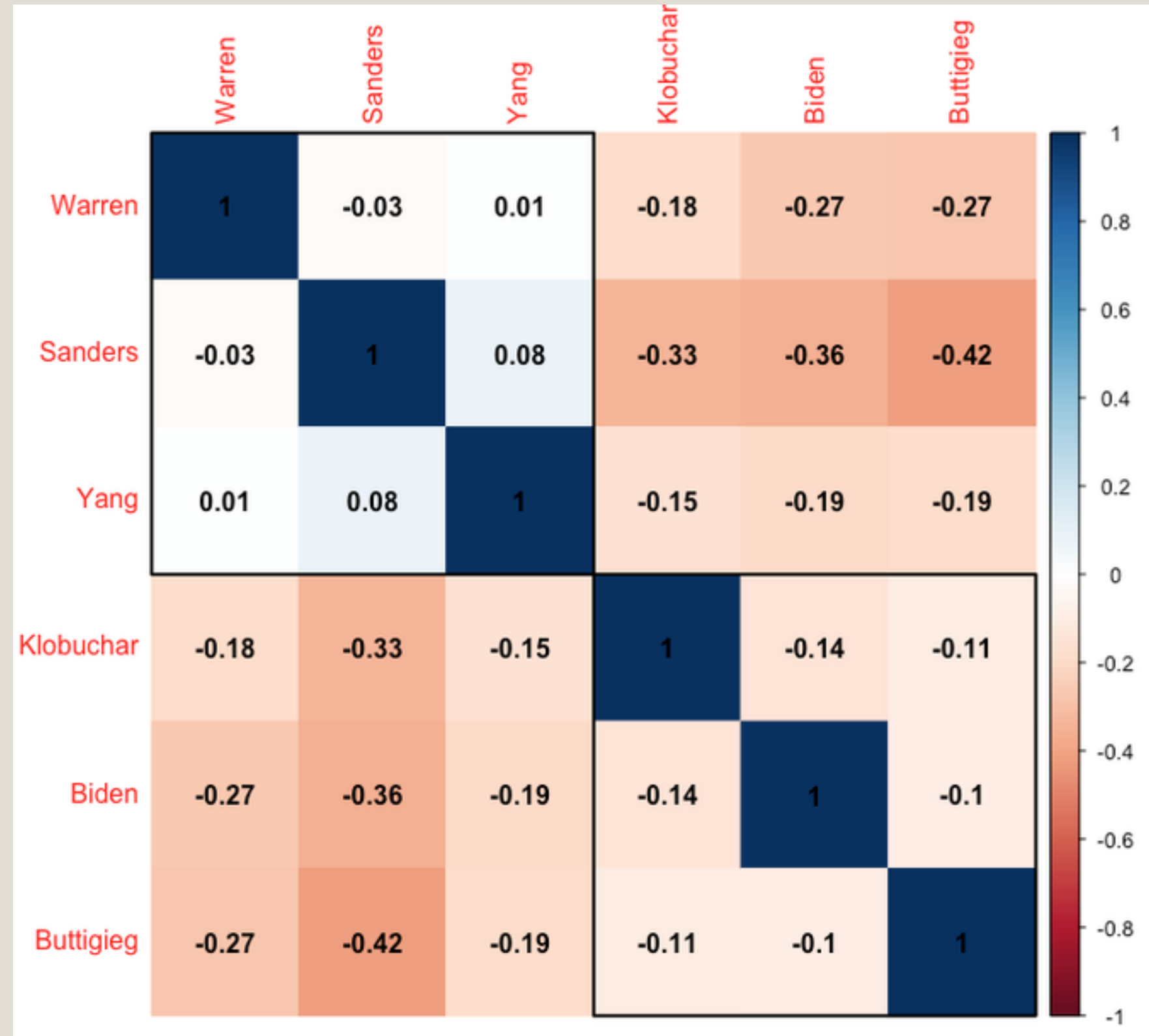
$$\hat{y} = a + bx_1 + cx_2 + \dots + dx_n$$

Bagımlı
Değişken

Regresyon
Sabit

Bagımsız
değişkenler

Korelasyon (Bağılılaşım) Matrisi



r (korelasyon katsayısı) nasıl yorumlanır

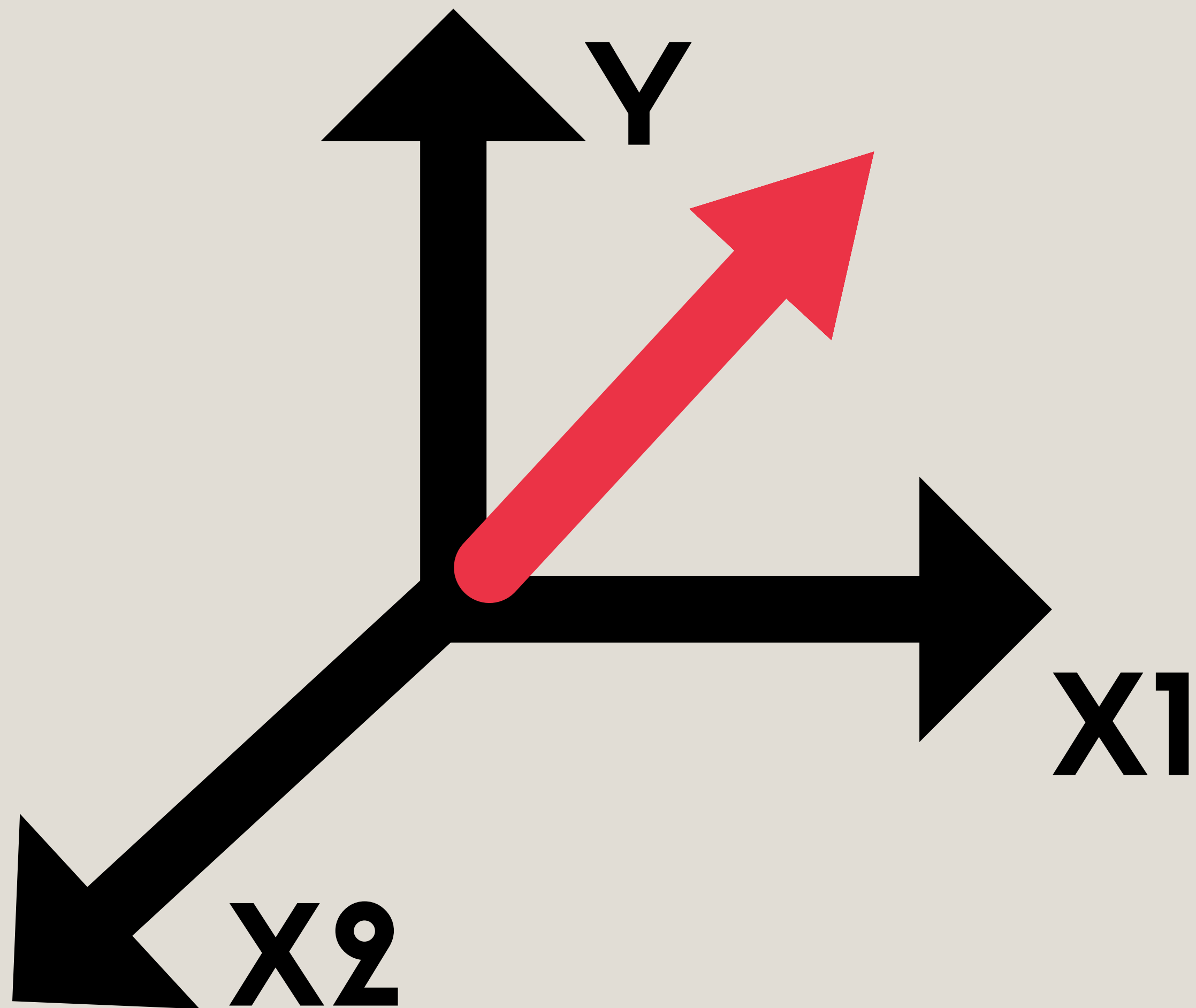
$r < |0.2|$ ise çok zayıf ilişki yada korelasyon yok.

$|0.2-0.4|$ arasında ise zayıf korelasyon.

$|0.4-0.6|$ arasında ise orta şiddette korelasyon.

$|0.6-0.8|$ arasında ise yüksek korelasyon.

$|0.8| > r$ ise çok yüksek korelasyon olduğu yorumu yapılır.



Python ile Çoklu lineer regresyon kodlaması