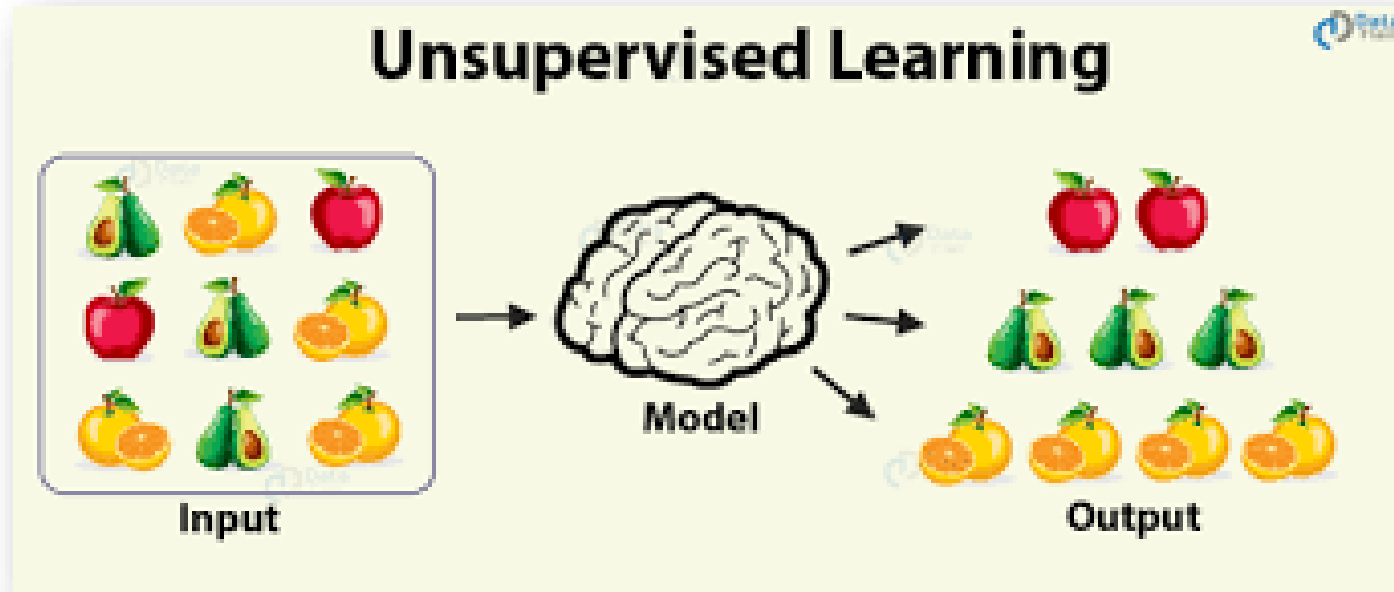


DENETİMSİZ ÖĞRENME



İÇİNDEKİLER

Makine Öğrenmesi Nedir?

Denetimli Öğrenme Nedir?

Denetimsiz Öğrenme Nedir?

Denetimli ve Denetimsiz Öğrenme Arasındaki Farklılıklar

Kümeleme Algoritması

1-) K-Means Kümeleme Algoritması

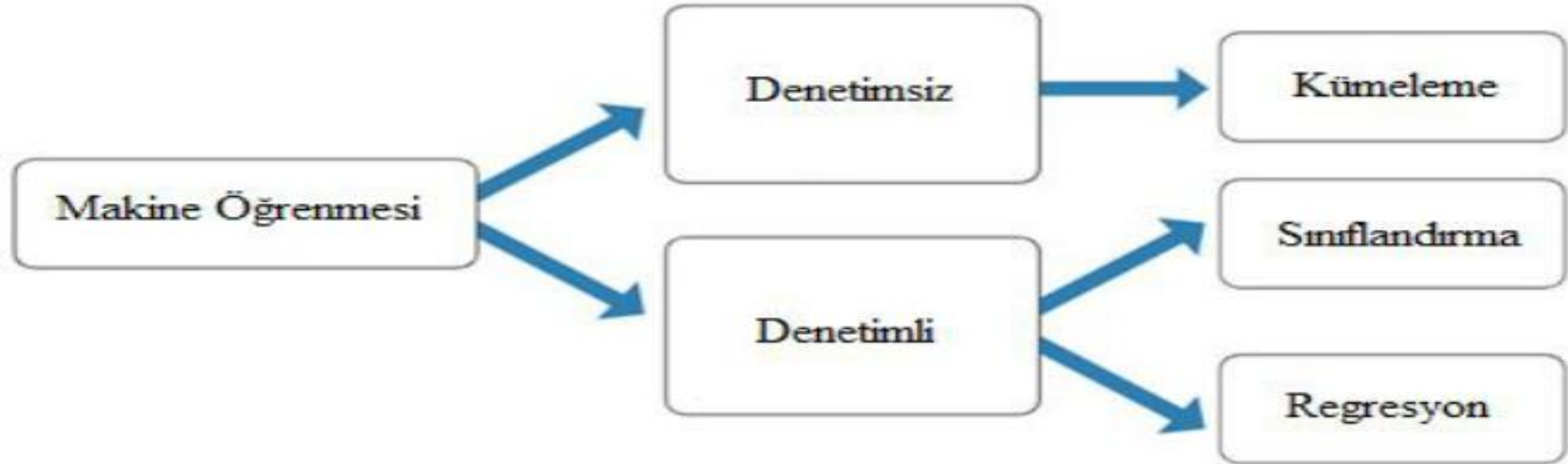
2-) Hiyerarşik Kümeleme Algoritması

Boyut Küçültme Algoritması

Pazar Sepeti Algoritması

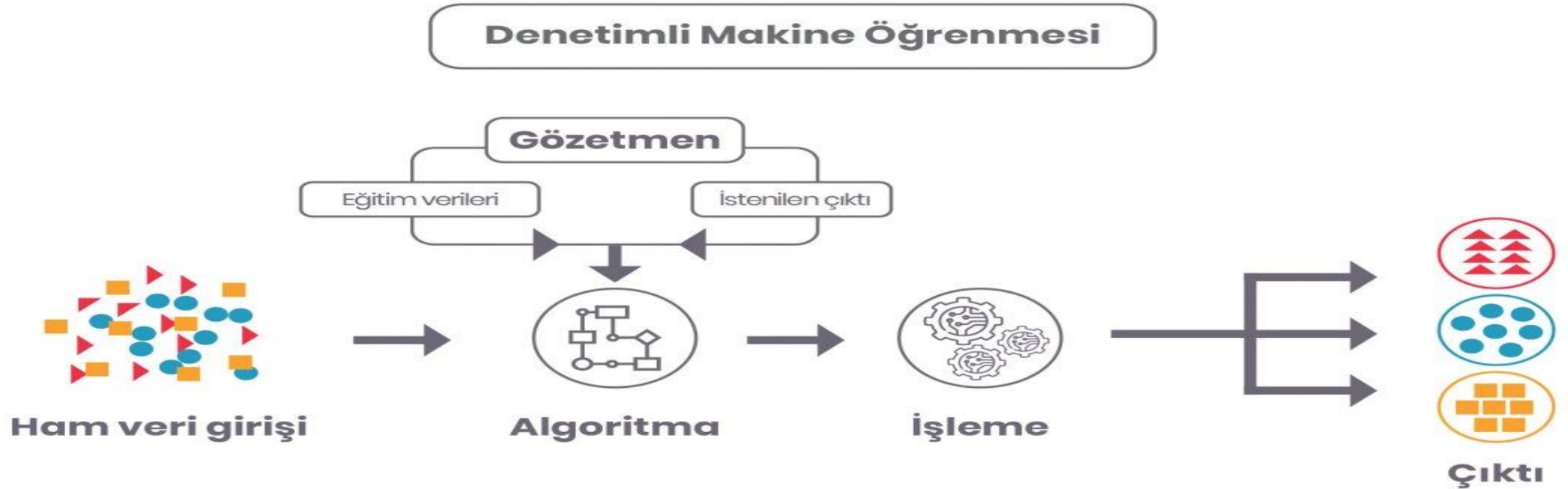
MAKİNE ÖĞRENMESİ NEDİR

Makine öğrenmesi, insanların öğrenme şekillerini taklit etmek için veri ve algoritmaların kullanımına odaklanıp doğruluğunu kademeli olarak artıran bir yapay zeka (AI) ve bilgisayar bilimi dalıdır.



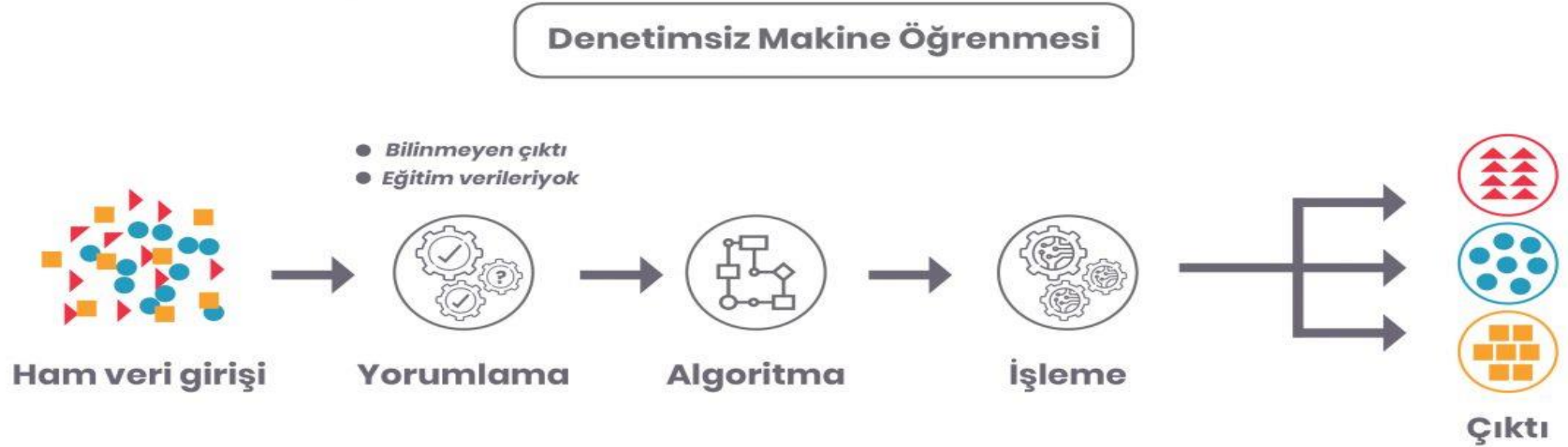
DENETİMLİ ÖĞRENME

Eğitim verilerinden, öngörülemeyen verileri tahmin etmenize yardımcı olmak için öğrenen bir algoritmadır. Denetimli makine öğrenimi, istenen sonuçları elde etmek için eğitim veri kümelerini kullanır.

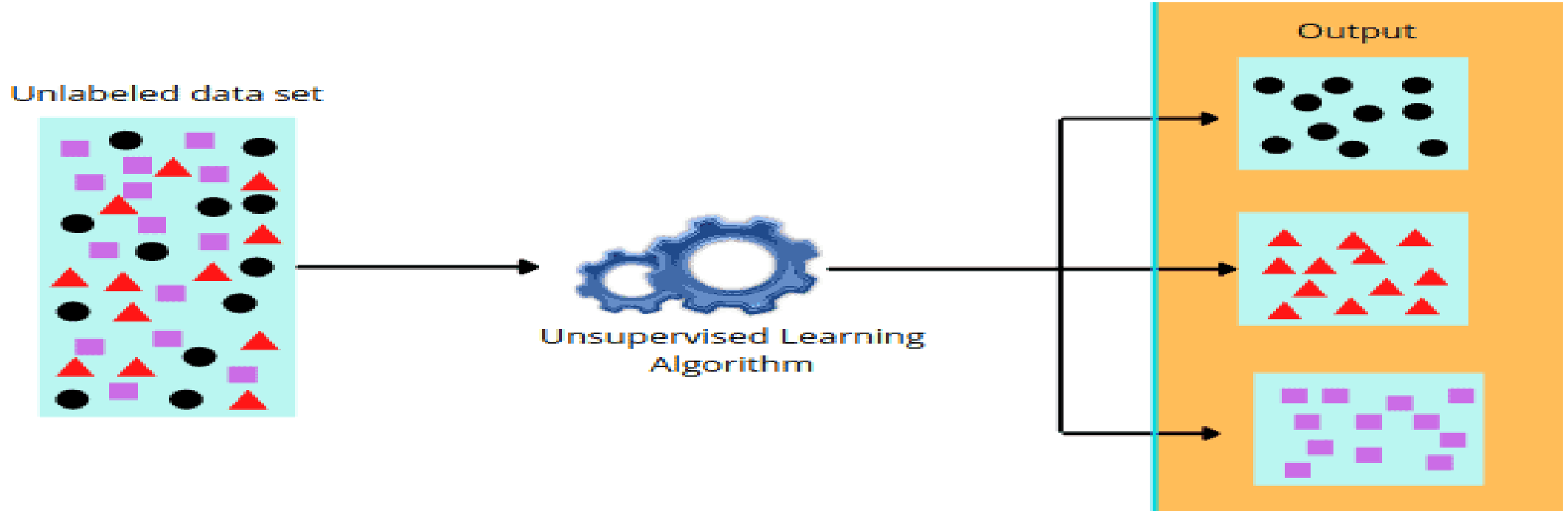


DENETİMSİZ ÖĞRENME

Denetimsiz Öğrenme etiketsiz veriler arasındaki gizli kalmış yapıyı/örüntüyü bulmaya çalışan bir makine öğrenmesi yöntemidir. Burada önemli olan elimizdeki verilerin etiketi hakkında bir bilginin olmadığı durumlarda denetimsiz öğrenme kullanılmaktadır.



- Denetimsiz öğrenmede elimizde sadece veriler vardır ve elimizdeki bu verilerden sonuç çıkarılmaya çalışılmaktadır. Aslında yapılan her işi kendi kendine öğreniyor
- ! Veriler etiketsiz olduğu için de Denetimsiz Öğrenme algoritması çıktısına kesinlikle doğrudur gözüyle bakmamak gerekir.

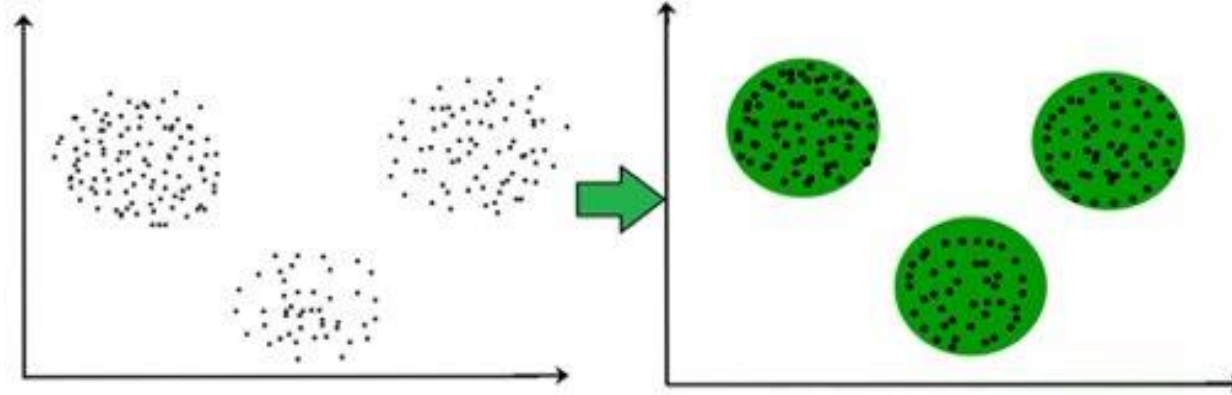


Denetimli ve Denetimsiz Öğrenme arasındaki farklılıklar

	Denetimli Öğrenme	Denetimsiz Öğrenme
Veri Tipi	Etiketli veriler	Etiketsiz veriler
Karmaşıklık	Az karmaşık	Daha karmaşık
Analiz	Çevrimdışı Analiz	Gerçek Zamanlı Analiz
Güvenilirlik	Doğru sonuçlar verir	Orta derecede güvenilir
Uygulama	Regresyon ve Sınıflandırma problemleri	Kümeleme ve ilişkilendirme problemleri

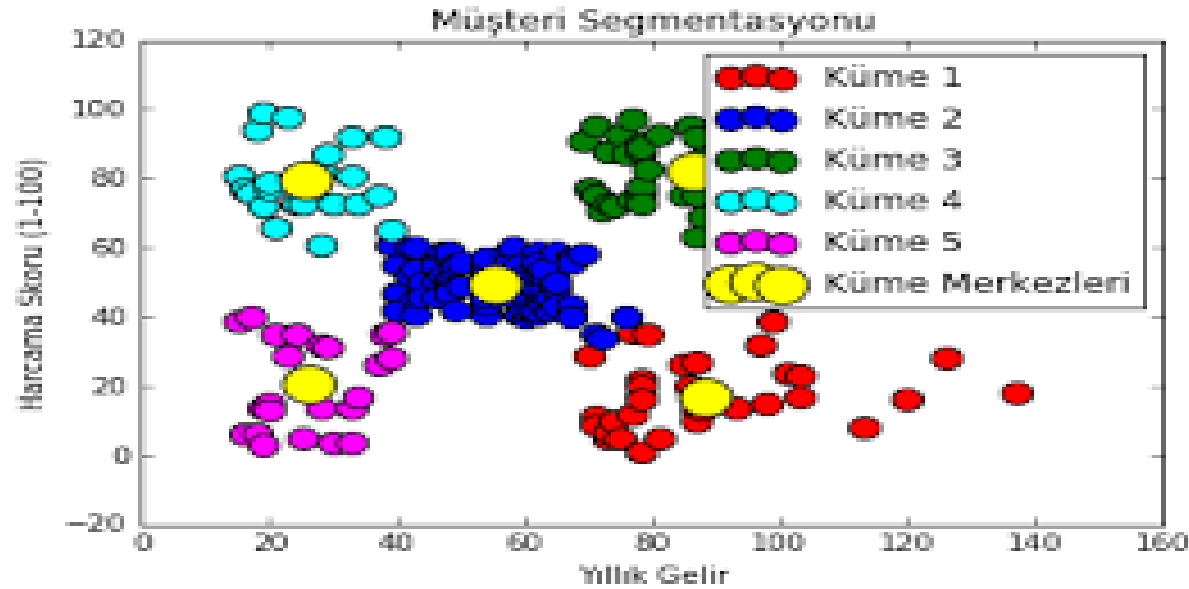
KÜMELEME ALGORİTMASI

- Veri kümesindeki elemanların yakınlık, uzaklık ve benzerlik gibi ölçütlere göre analiz edilerek kendi arasında gruptandırmaya kümeleme denir.
- Temelde veriler kendi aralarındaki benzerlik ve farklılıklar analiz edilerek gruplandırılır.
- Kümeleme algoritmalarında kaç küme oluşması gerektiğine kullanıcı veya algoritma karar verir. Küme sayısı ile ayrıntı düzeyi belirlenir.



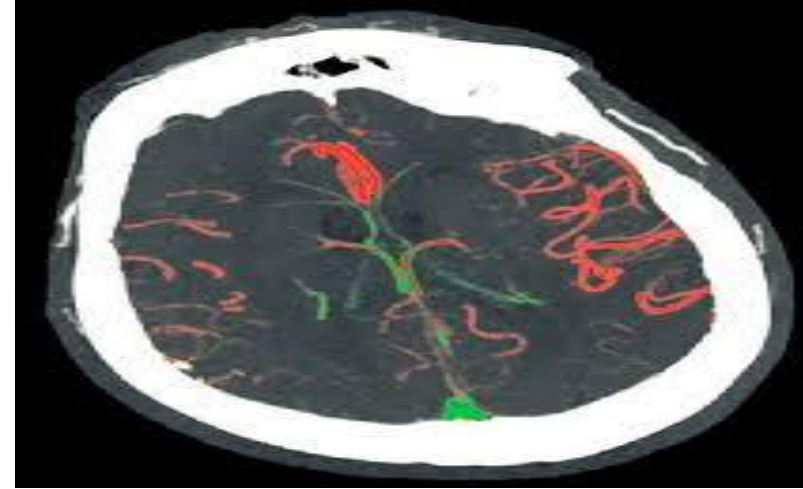
Kümeleme Algoritması Kullanım Alanları

- **1-) Müşteri Segmentasyonu** : Müşterilerin verileri ön tanımsız olarak makineye verilir. Makine ortak özelliklere göre sınıflandırma yapar. Müşterilerin geçmiş verilerine bakılarak veya kendisine benzer kişilerin davranışlarına bakılarak ürün tavsiye edilir. Kişilere özel kampanyalar yapılır.



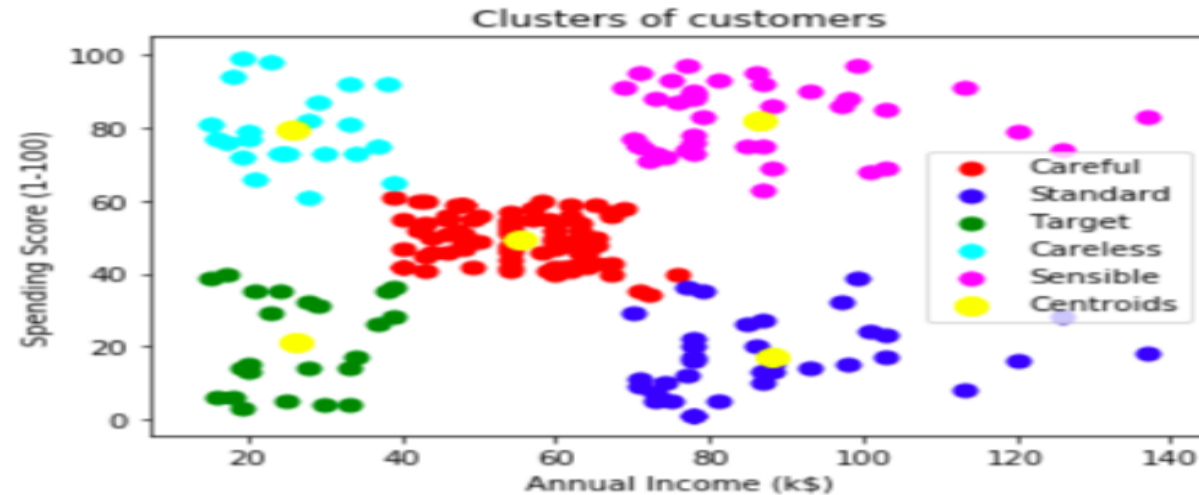
Kümeleme Algoritması Kullanım Alanları

- **2-) Pazar Segmentasyonu :** Pazar araştırması kapsamında yaş grupları, kazanç dilimleri, kentsel, kırsal veya banliyö konumu gibi kategorileri belirlemek için kullanılabilir. Pazarlamada, farklı müşteri gruplarının en alakalı mesajlarla hedeflenebilmesinde küme analizi kullanılır. Satın alma kalıplarına göre müşteri gruplarını karakterize edilir.
- **3-) Sağlıkta Kümeleme :** Özellikle görüntü işlemede kümeleme çok kullanılır. Makine görüntüyü alıp analiz eder ve ardından kümeleme yapar.



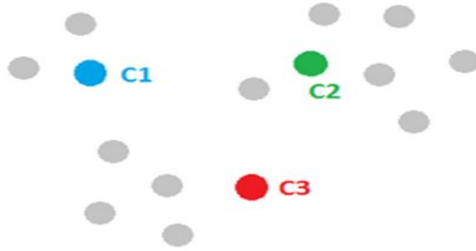
1-) K-Means Kümeleme Algoritması

- K-Means kümeleme algoritması yinelemeli bir kümeleme algoritmasıdır.
- "K" değeri kaç kümeye ayıracağımızı belirleyen faktördür.
- K değeri her yinelemede en yüksek değeri bulmaya yardım eder.
- Başlangıçta K bizim tarafımızdan belirlenebilir. Sonrasında seçilen K değerine göre veriler K tane merkez noktada kümelenir.
- Kararlı hale gelinceye kadar merkez bulma döngüsü devam eder.

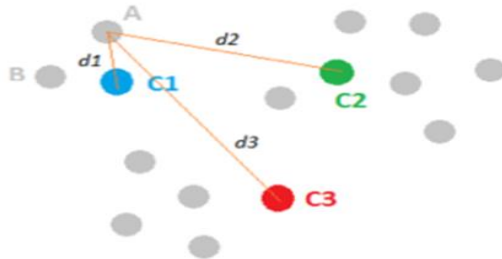


1-) K-Means Kümeleme Algoritması Adımları

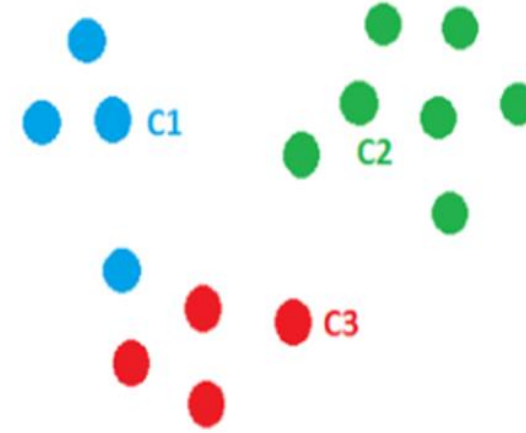
- Adım 1 : Kümeler için bir merkez noktası rastgele seçilir.



- Adım 2: Merkez eleman dışındakiler en yakın merkez noktaya göre kümelendir. Bunun için her bir elemanın merkez noktalara olan uzaklıkları ölçülür.

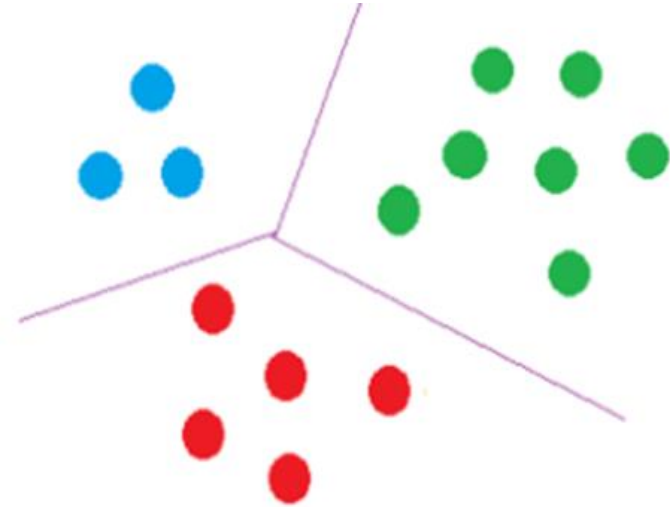
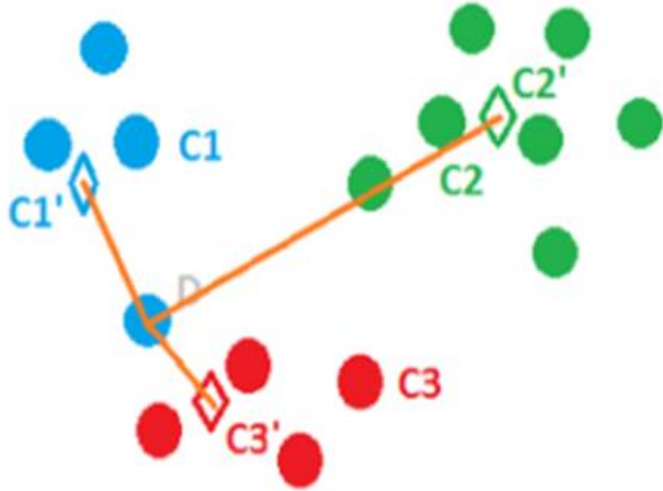


- Adım 3 : Kümelenen elemanlarımız şöyle bir hal alır.



1-) K-Means Kümeleme Algoritması Adımları

- Adım 4 : Ardından aynı kümeler içerisinde tekrar bir merkez noktası hesaplanır.
- Adım 5 : Daha sonra merkez noktası kararlı hale gelinceye kadar ilk iki adım tekrarlanır.



1-) K-Means Kümeleme Algoritması

- K-Means algoritmasının bazı problemleri vardır. Yapılan örnekte küme sayısını üç olarak aldık. Fakat küme sayısını kullanıcı belirleyeceği için bu sayı her zaman optimum olmayabilir.
- Başka bir problem ise başlangıçta rastgele olarak atanan merkez noktaları her zaman başarılı bir kümeleme işlemi başlatmayabilir. Çünkü algoritmanın gidişatı başlangıçta rastgele olarak atanan bu merkez noktalarına göre şekillenmektedir. Başlangıçta atanan bu rastgele noktalar için K-Means++ algoritması mevcuttur.
- Optimum küme sayısını belirlemek için de WCSS (within-cluster sums of squares) yöntemi kullanılmaktadır.

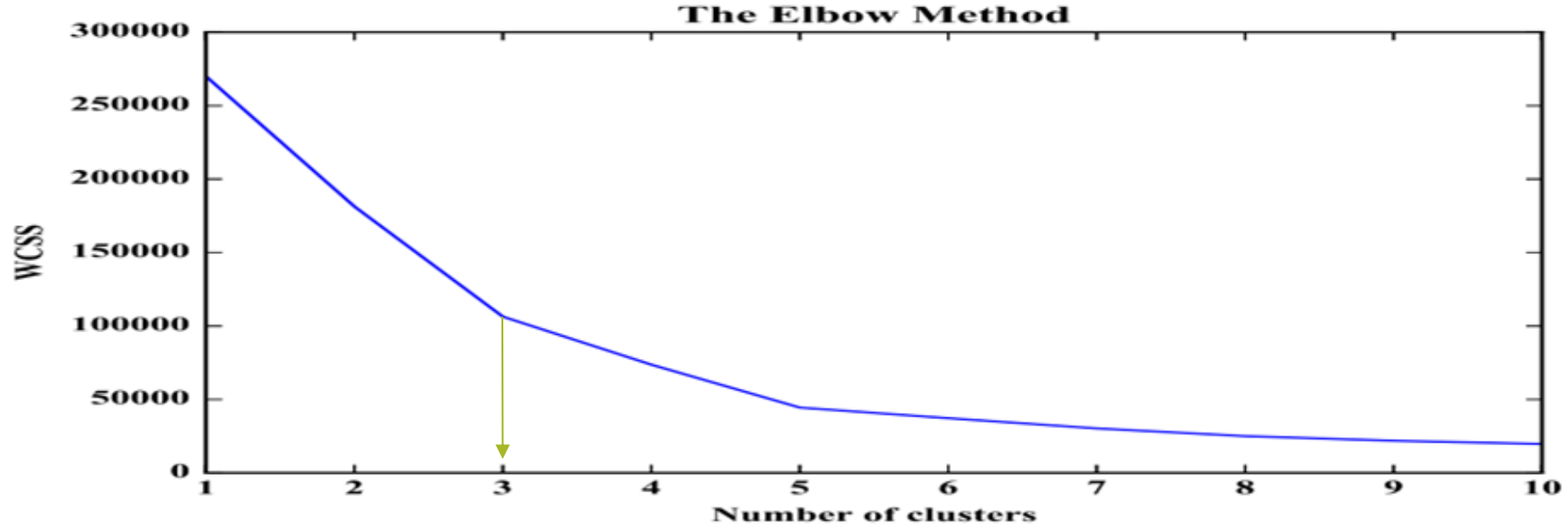
1-) K-Means Kümeleme Algoritması

- Amaç her bir küme için o kümedeki noktaların merkez noktalarına olan uzaklıklarının karelerini hesaplamak.
- Kümeleme algoritmasında küme elemanları arasındaki mesafe minimum, her küme arasındaki mesafe ise maksimum ise o küme başarılı bir kümedir.

$$WCSS = \sum_{P_i \text{ in Cluster 1}} \text{distance}(P_i, C_1)^2 + \sum_{P_i \text{ in Cluster 2}} \text{distance}(P_i, C_2)^2 + \sum_{P_i \text{ in Cluster 3}} \text{distance}(P_i, C_3)^2$$

1-) K-Means Kümeleme Algoritması

- Eğer küme sayısını olabildiğince artırıp küme içindeki her bir eleman arası mesafeyi azaltırsak WCSS değeri o kadar azalır ve başarılı bir kümeleme gerçekleştirmiş oluruz. Fakat küme sayısını artırmanın da bir limiti vardır. Eğer küme sayısını her bir eleman ayrı bir küme olacak şekilde yaparsak küme elemanları arasındaki mesafe 0 olur ve WCSS de 0 olur. Fakat bu durumda modelimiz overfitting yani ezberleme durumuna uğrar. Onun için optimum küme sayısını WCSS grafiğindeki ani eğim değişimi olan yere göre belirlemeliyiz. Bu grafikteki en uygun değer 3'tür.



1-) K-Means Kümeleme Algoritması Kullanım Alanları

- Belge Sınıflandırılması : Belgeleri etiketlere, konulara ve belgenin içeriğine göre birden fazla kategoride kümeleyin. Bu standart sınıflandırma problemi K-Means ile çözülebilir.
- Suç Yerlerinin Belirlenmesi : Bir şehirdeki belirli bölgelerde mevcut olan suçlarla ilgili veriler, suç kategorisi, suç alanı ve ikisi arasındaki ilişki, bir şehirdeki ya da bölgedeki suça eğilimli alanlara ilişkin kaliteli bilgiler verebilir.
- BT Uyarılarının Otomatik Kümelenmesi : Ağ, depolama veya veritabanı gibi büyük BT altyapı teknoloji bileşenleri büyük hacimli uyarı mesajları içerir. Uyarı mesajları potansiyel olarak operasyonel sorunlara işaret ettiğinden sonraki işlemler için önceliklendirmenin manuel olarak taranmaları gerekir. Verilerin kümelenmesi uyarı kategorileri hakkında bilgi verebilir. Böylece ortalama onarım süresinin azalmasına tahminlerde yardımcı olarak katkı sağlar.
- Görüntü Tanıma
- Oyuncu Analizi
- Çağrı Kaydı Detay Analizi
- Dolandırıcılık Tespiti

2-) Hiyerarşik Kümeleme Algoritması Nedir?

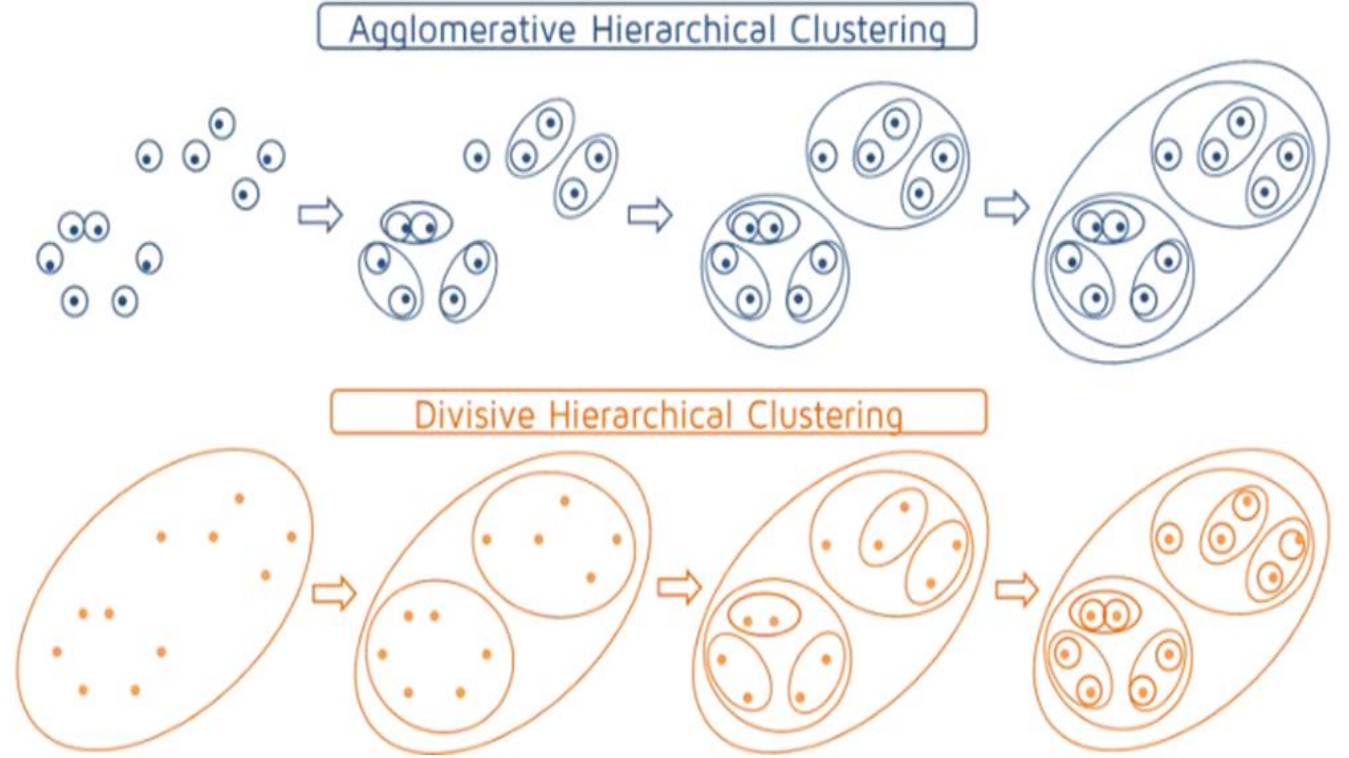
- Hiyerarşik kümeleme, birleştirici ve ayrıştırıcı hiyerarşik kümeleme yöntemleri olmak üzere iki türe sahiptir. Birleştirici hiyerarşik kümelemede başlangıçta her bir değer bağımsız bir küme olarak değerlendirilir ve bu değerler çeşitli algoritmalarla birleştirilip her aşamada bir üst küme oluşturulur. Ayrıştırıcı kümelemede ise başlangıçta tüm değerler bir küme olarak değerlendirilip yine çeşitli ayrıştırma algoritmalarıyla alt kümeler elde edilir. İşlem sonunda **dendrogram** denen tüm nesnelerin ilişkilerini gösteren bir ağaç yapısı oluşur.

Neden Hiyerarşik Kümeleme Algoritması?

- K-means algoritmasının en kötü taraflarından biri küme sayısına karar vermenin zorluğudur. Küme sayısına ancak sistemi çok iyi bilen biri karar verebilir ki bu da probleme göre değişir. Örneğin bir ödev verdiniz ve hangi öğrencinin hangi öğrenciye yakın ödev yaptığını ve kaç küme oluştuğunu merak ettiğinizde küme sayısını önceden bilmenize olanak yoktur. Bir ödevde 3 küme çıkar, diğer ödevde 20 küme çıkabilir. İşte bu noktada x-means ve single pass gibi algoritmalar olsa da hiyerarşik kümeleme küme oluşturma için esnek bir yapı olması sebebiyle bir adım öne çıkar.
- Gözetimsiz öğrenme biri olan Hiyerarşik kümeleme, diğer kümeleme yöntemlerine göre daha bilgilendirici bir çıktı üretir. Hiyerarşik kümelemede, küme sayısı önceden belirlenmesi gerekmemektedir.

Bu yöntemde ise iki adet yaklaşım bulunmakta:

- **Agglomerative:** Her bir veri öncelikle kendine ait bir küme oluşturur. Ardından birbirine en yakın olan iki küme birleşerek yeni bir küme oluşturur ve bu işlem tek bir büyük küme oluşuncaya dek devam eder.
- **Divisive:** Bu yaklaşımda ise tüm veriler öncelikle tek bir küme içerisinde ele alındıktan sonra her bir veri noktası birer küme oluncaya dek bölünme işlemi gerçekleşiyor.



Divisive'e Devam Edelim:

Ayrıştırıcı hiyerarşik kümeleme yöntemin adımları ve akış şeması:

- Her nesneler için değer oluştur ve her nesneyi bir küme olarak ata. (Bu durumda N tane nesne için N adet küme oluşur.) $N \times N$ bir matris oluştur.
- En yakın olan iki nesneyi grupla. Matrisi kullanarak nesneler arası uzaklık değerlerini hesapla.
- $N=N-1$ boyutunda yeni matrisi oluştur. Taki N değeri 1 oluncaya kadar Adım 2. ye geri dön.

Adım 1:

Verinin Hazırlaması ve matrisin oluşturulması

- Veri hazırlama aşaması çalışma alanına göre farklılık gösterebilir. Bu kümeleme yönteminde veriler arasındaki benzerlikleri içeren bir matris oluşturmamız gerekir. Örneğin gönderilen ödevleri için kelime veya karakter benzerliği kullanılabilir. Dokümanlar arasındaki benzerlikler bulunduğundan sonra aşağıdaki gibi bir matris elde edebiliriz.

	1	2	3	4	5	6
1	0.000	0.458	0.177	0.419	0.4	0.161
2	0.458	0.000	0.486	0.263	0.263	0.462
3	0.177	0.486	0.000	0.000	0.300	0.004
4	0.419	0.263	0.300	0.000	0.000	0.276
5	0.419	0.263	0.300	0.000	0.000	0.276
6	0.161	0.462	0.004	0.276	0.276	0.000

	1	2	3	4	5	6
1						
2	0.458					
3	0.177	0.486				
4	0.419	0.263	0.300			
5	0.419	0.263	0.300	0.000		
6	0.161	0.462	0.004	0.276	0.276	

- Birinci matris ödevler ve ödevler arasındaki benzerlikleri göstermektedir. Örneğin 1 ve 2 nolu öğrenciler 0.458 oranında bir benzeşim bulunmaktadır. Tabii, 2 – 1 nolu bakışta aynıdır, yani simetriktir ve 1-1 nolu dokümanlar ise beklenildiği aynı dokümanlar olduğu için hiç bakmaya gerek yoktur. Bu noktada sadece matris-2'deki verileri kullanılabilir. (iç içe olan for döngüsü üzerinde bir düzenleme ile veya bir if şartı ile gereksiz ver benzerlikleri kolayca göz ardı edilebilir.)

Adım 2:

- Matris üzerinde en küçük değer (maksimum benzerlik) 4 ve 5. öğrenci ödevlerinin benzerlik oranını gösteren o değeridir. Başka bir deyişle bu iki öğrencinin ödevleri tamamen aynıdır. 4 ve 5'ten yeni bir küme üretilebilir. Şimdi birleştirme işlemini yapalım ve yeni matrisimizi oluşturalım.

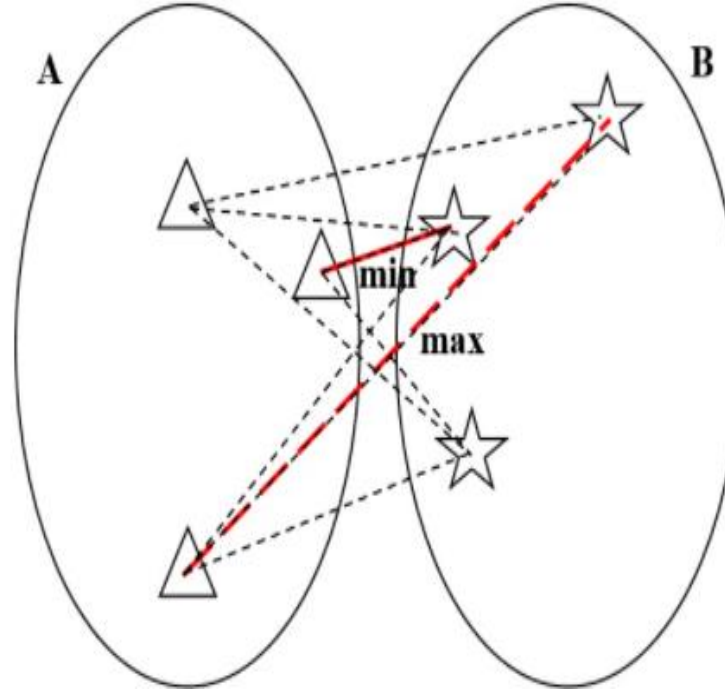
	1	2	3	4, 5	6
1					
2	0.458				
3	0.177	0.486			
4,5	???	???	???	????	
6	0.161	0.462	0.004	???	

???: Algoritma tarafından belirlenecek değerler

4 ve 5. kişiler birinci kümemiz oldu ve birleşti. "???" li kısımların hesabını yapalım. Bu noktada "???" kısımlar farklı şekillerde hesaplanabilir.

Hiyerarşik Kümeleme Hesaplama Yöntemleri: Tek Bağlantı (Single-Link), Tam Bağlantı (Complete-Link) ve Ortalama Bağlantı (Average-Link) Kriterleri

- Hiyerarşik kümelemede, sayısal değerler üzerinden farklı hesaplama yöntemlerine göre bir küme üretilir. Bu noktada üç farklı Tek Bağlantı, Tam Bağlantı ve Ortalama Grup yöntemlerinden biri kullanılabilir.
- Tek bağlantı kümeleme, her adımda en küçük mesafeye sahip en yakın üyeler kullanarak iki kümenin birleştirilmesidir. Bu algoritmaya, en yakın komşu algoritması da denir. Tam bağlantı kümeleme, tek bağlantının tam tersine her adımda en büyük mesafeye sahip en uzak üyeler kullanarak iki kümenin birleştirilmesidir. Bu algoritmaya bu sebepten dolayı en uzak komşu algoritması da denir. Ortalama grup algoritması ise iki karşılıklı küme arasındaki tüm mesafelerin ortalamasıdır. Bu algoritma ile tek bağlantı ve tam bağlantı algoritmalarında elde edilen değerlerin arasında bir değer elde edilir (Manning ve ark., 2008).



Tek Bağlantı:

$$d_{A \rightarrow B} = \min(d_{i,j})$$

$\forall i \in A$
 $\forall j \in B$

Tam Bağlantı:

$$d_{A \rightarrow B} = \max(d_{i,j})$$

$\forall i \in A$
 $\forall j \in B$

Ortalama Grup:

$$d_{A \rightarrow B} = \text{avg}(d_{i,j})$$

$\forall i \in A$
 $\forall j \in B$

- (4, 5) kümesi yapıldıktan sonra yeni değerler hesaplanmalıdır. Bu hesap sırasında Şekil 1'deki matris bilgileri kullanılır. Bu aşamada min, max veya avg algoritmalarından biri kullanılır. Bu örnekte min değerini kullanılacaktır. Min yani Tek Bağlantı hesabı:

$$d_{(4,5) \rightarrow 1} = \min(0.419, 0.419), \quad d_{(4,5) \rightarrow 2} = \min(0.263, 0.263), \quad d_{(4,5) \rightarrow 3} = \min(0.300, 0.300), \\ d_{(4,5) \rightarrow 6} = \min(0.276, 0.276)$$

Tek bağlantıya göre bu kümelerden en küçük değerler seçilir. İki ödev birbiriyle aynı olduğu için tüm benzerlikler aynı çıkmıştır. Şimdi ikinci döngüye geçelim. Yeni oluşan matriste en küçük değer 0.004'tür. Başka bir deyişle 3 ve 6'ncı öğrencilerden yeni bir küme oluşturulacaktır.

2. Döngü Başlangıç

	1	2	3	4, 5	6
1					
2	0.458				
3	0.177	0.486			
4, 5	0.419	0.263	0.300		
6	0.161	0.462	0.004	0.276	

2. Döngü Sonu

	1	2	3,6	4, 5
1				
2	0.458			
3, 6	0.161	0.462		
4, 5	0.419	0.263	0.276	

Adım

3:

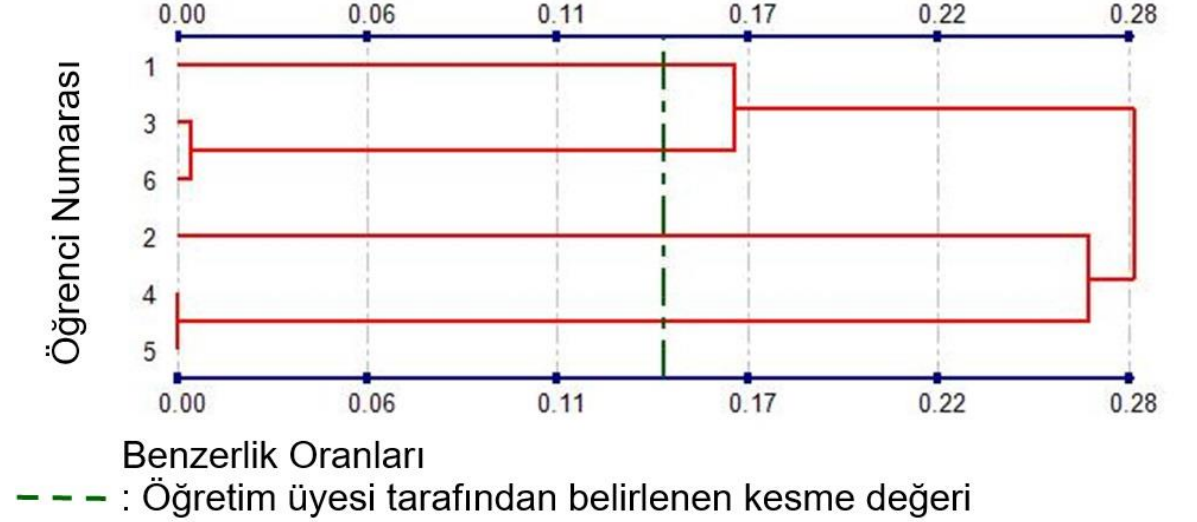
- Her döngünün sonunda küme sayısı 1 azalır ve küme sayısı 1 oluncaya kadar aynı işlemler tekrar eder. Her adım sonunda bir birleşim işlemi olur. Adım adım oluşan kümeler / ilişkileri ve seçilen minimum değerler aşağıda verilmiştir.

Adım	Kümeler	Minimum
0.	1, 2, 3, 4, 5, 6	-
1.	4, 5	0.000
2.	3, 6	0.004
3.	1, (3, 6)	0.161
4.	2, (4, 5)	0.263
5.	(1, (3, 6)), (2, (4, 5))	0.276

- Bu aşamada küme sayısına karar vermek gerekir. Küme sayısı karar vermek için dendrogram kullanılır.

Dendrogram

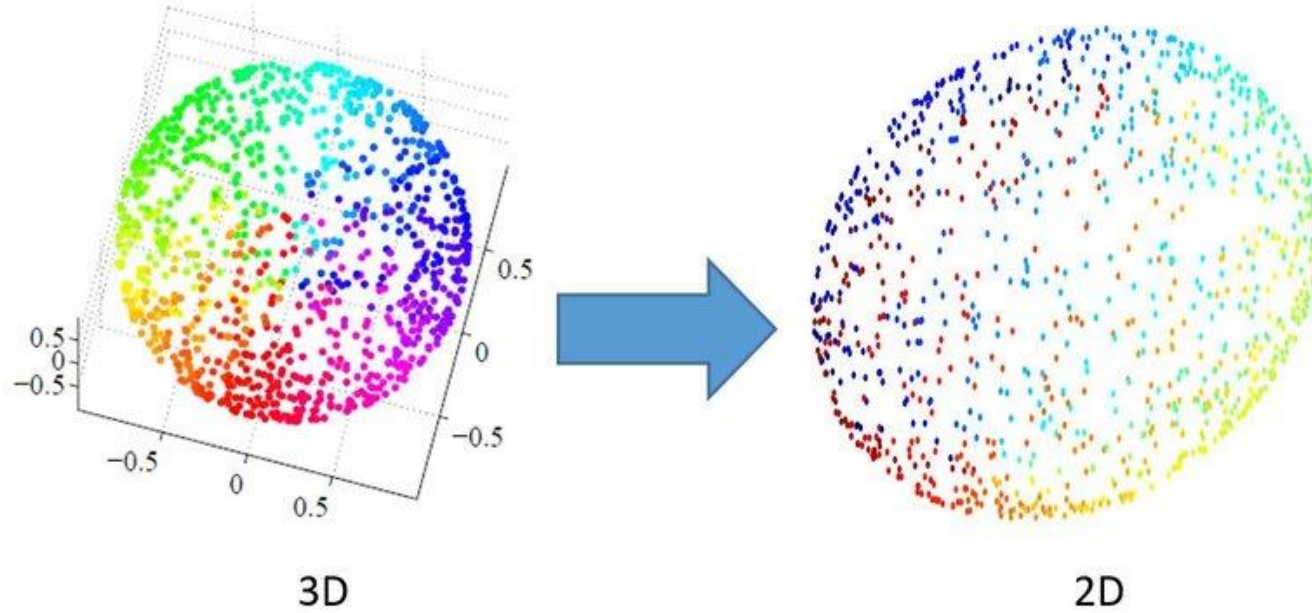
- Dendrogram tüm nesnelerin ilişkilerini gösteren bir ağaç yapısı oluşur. Bu grafik, küme sayısını belirlemede kolaylık sağlar. Küme sayısı belirlemek için kesme değeri belirlenmelidir. Ödevlerde benzerlik kaçınılmazdır.



Öncelikle (4, 5)'in 0 değeri çizilmiş ardından (3, 6)'nın 0.004 değeri çizilmiştir. Bu şekilde tüm kümeleme sonuçlarının çizimini yapılır. Ödev değerlendirmede dendrogram sayesinde öğretim üyesi çok rahat şekilde yapılan ödevleri yorumlayabilir. Örneğin 4 ve 5 nolu öğrencilerin aynı ödevi gönderdiği 3 ile 6 nolu öğrencilerin ise çok az bir farklılık içerdiğini görmektedir. 1 nolu öğrencinin 3 ve 6 nolu öğrencilere benzer bir ödev yaptığı görülmektedir. Bu benzerlik öğretim üyesi tarafından kopya oluşturacağı düşünülürse 0.161 büyük bir kesme değeri seçilebilir. Eğer farklı bir ödev yaptığı düşüncesinde ise şekildeki gibi 0.15 kesme değerini seçebilir. Bu kesme değeri kullanıldığında 1, (3, 6), 2, (4, 5) olmak üzere 4 adet küme oluşmaktadır. 1 ve 2 nolu öğrenciler orjinal ödev yaptığı görülmektedir.

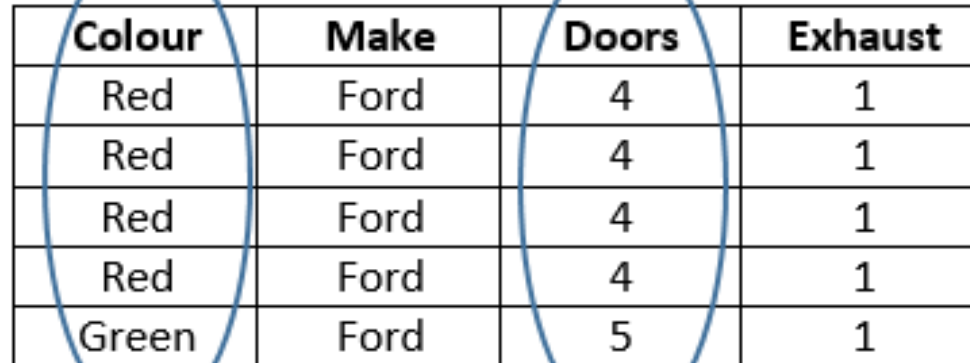
Boyut Küçültme Nedir?

- Gerçek hayattaki veriler çok fazla boyuta (özniteliğe) sahip oluyor ve boyut büyüdükçe veri temizlemeden model kurmaya bütün süreçlerde harcamamız gereken **zaman ve kaynaklar** artıyor. Boyut azatma bu sorunun önüne geçmek için kullanılan yöntemlerden biridir. Hemen her veri setinde bazı öznitelikler arasında yüksek korelasyon oluyor ve bu bizim gereksiz bilgiye sahip olmamıza ve modelimizde **overfitting problemine** sebep olabiliyor.



Nasıl Uygulanır ?

- Boyut küçültmenin en kolay yolu verimizi en iyi tanımlayan öznitelikleri bulup diğerlerini atmaktır (**öznitelik seçimi — feature selection**). Dikkat edilmesi gereken nokta en az bilgi kaybıyla bu işi yapmaktır ve aslında önemli olan öznitelikleri silmemektir. Bunu yapmak için verideki dağılımın maksimum varyansını-bilgisini tutan minimum sayıda değişken oluşturuyoruz. Eğer bir değişken her örnek için aynı değere sahip ise gereksiz bir değişkendir. Biz en yüksek varyansa sahip olan değişkenleri bulmalıyız.



Colour	Make	Doors	Exhaust
Red	Ford	4	1
Red	Ford	4	1
Red	Ford	4	1
Red	Ford	4	1
Green	Ford	5	1

HDI(İnsani Kalkınma Endeksi)

- Buna en güzel örneklerden biri HDI(İnsani Kalkınma Endeksi) hesaplaması olabilir. Ülkelerin sahip olduğu değişkenlerden bazıları (wiki): resmi dil, yüz ölçümü (toplam), su oranı, nüfus, nüfus yoğunluğu, GSYH, para birimi, trafik akışı, hukuk, dış ilişkiler, din, eğitim, sağlık, sanayi, tarım, turizm... Bunlara ek olarak ülkeden ülkeye değişen yüzlerce değişken sıralayabiliriz.

HDI(İnsani Kalkınma Endeksi)

- PCA metodu sanayi, tarım, turizm vs. ekonomiyle ilgili yüzlerce değişkeni ve nüfusu kullanarak kişi başına düşen gayri safi milli gelir diye tek bir değişken oluşturuyor. Yani yüzlerce boyutluk bilgi en az bilgi kaybıyla tek boyutta tutulmuş oluyor.
- HDI hesaplamalarında sadece 5 öz niteliğe bakılarak pek de itiraz edilmeyecek bir tablo karşımıza çıkıyor.
- Kişi başına düşen gayri safi milli gelir
- İnsan gelişmişlik endeksi — HDI sıralaması
- Beklenen ortalama yaşam süresi
- Beklenen eğitim yılı
- Ortalama eğitim yılı

		Human Development Index (HDI)	Life expectancy at birth	Expected years of schooling		Mean years of schooling		Gross national income (GNI) per capita		GNI per capita rank minus HDI rank	HDI rank
HDI rank	Country	Value	(years)	(years)		(years)		(2011 PPP \$)		2015	2014
		2015	2015	2015	a	2015	a	2015		2015	2014
	VERY HIGH HUMAN DEVELOPMENT										
1	Norway	0.949	81.7	17.7		12.7		67,614		5	1
2	Australia	0.939	82.5	20.4	b	13.2		42,822		19	3
2	Switzerland	0.939	83.1	16.0		13.4		56,364		7	2
4	Germany	0.926	81.1	17.1		13.2	c	45,000		13	4
5	Denmark	0.925	80.4	19.2	b	12.7		44,519		13	6
5	Singapore	0.925	83.2	15.4	d	11.6		78,162	e	-3	4
7	Netherlands	0.924	81.7	18.1	b	11.9		46,326		8	6
8	Ireland	0.923	81.1	18.6	b	12.3		43,798		11	8
9	Iceland	0.921	82.7	19.0	b	12.2	c	37,065		20	9
10	Canada	0.920	82.2	16.3		13.1	f	42,582		12	9
10	United States	0.920	79.2	16.5		13.2		53,245		1	11
12	Hong Kong, China (SAR)	0.917	84.2	15.7		11.6		54,265		-2	12
13	New Zealand	0.915	82.0	19.2	b	12.5		32,870		20	13
14	Sweden	0.913	82.3	16.1		12.3		46,251		2	15
15	Liechtenstein	0.912	80.2	14.6	g	12.4	h	75,065	a,i	-11	14
16	United Kingdom	0.909	80.8	16.3		13.3		37,931		10	16
17	Japan	0.903	83.7	15.3		12.5	c	37,268		10	17
18	Korea (Republic of)	0.901	82.1	16.6		12.2		34,541		12	18
19	Israel	0.899	82.6	16.0		12.8		31,215		16	19
20	Luxembourg	0.898	81.9	13.9		12.0		62,471		-12	20

PCA Metodu

- Boyut azaltma işlemlerinde sıklıkla kullanılan PCA verideki gerekli bilgileri ortaya çıkarmada oldukça etkili bir yöntemdir. PCA'nın arkasında yatan temel mantık çok boyutlu bir veriyi, verideki temel özellikleri yakalayarak daha az sayıda değişkenle göstermektir. En iyi sonucu elde etmek için genelde öncelikle standardizasyon yapılır.
- `x = StandardScaler().fit_transform(x)`

					Standardization ➡					
sepal length	sepal width	petal length	petal width	sepal length		sepal width	petal length	petal width		
0	5.1	3.5	1.4	0.2		0	-0.900681	1.032057	-1.341272	-1.312977
1	4.9	3.0	1.4	0.2		1	-1.143017	-0.124958	-1.341272	-1.312977
2	4.7	3.2	1.3	0.2		2	-1.385353	0.337848	-1.398138	-1.312977
3	4.6	3.1	1.5	0.2		3	-1.506521	0.106445	-1.284407	-1.312977
4	5.0	3.6	1.4	0.2		4	-1.021849	1.263460	-1.341272	-1.312977

PCA Metodu

- Sonrasında PCA uygulanılır.
- ```
from sklearn.decomposition import PCA
pca = PCA(n_components=2)
principalComponents = pca.fit_transform(x)
principalDf = pd.DataFrame(data = principalComponents, columns = ['principal component 1', 'principal component 2'])
```

|   | sepal length | sepal width | petal length | petal width |                            | principal component 1 | principal component 2 |
|---|--------------|-------------|--------------|-------------|----------------------------|-----------------------|-----------------------|
| 0 | -0.900681    | 1.032057    | -1.341272    | -1.312977   | PCA<br>(2 components)<br>→ | -2.264542             | 0.505704              |
| 1 | -1.143017    | -0.124958   | -1.341272    | -1.312977   |                            | -2.086426             | -0.655405             |
| 2 | -1.385353    | 0.337848    | -1.398138    | -1.312977   |                            | -2.367950             | -0.318477             |
| 3 | -1.506521    | 0.106445    | -1.284407    | -1.312977   |                            | -2.304197             | -0.575368             |
| 4 | -1.021849    | 1.263460    | -1.341272    | -1.312977   |                            | -2.388777             | 0.674767              |

# PAZAR SEPETİ ANALİZİ (MARKET BASKET ANALYSIS)

Pazar Sepeti Analizi Nedir?

Pazar Sepeti Uygulamaları

Apriori Algoritması kullanarak Pazar Sepeti Analizi

Pazar Sepeti Örnekler

# Pazar Sepeti Analizi Nedir?

- Pazar sepeti analizi, tüketicilerin satın alma davranışlarının analiz ederek ürünler arasındaki ilişkileri ortaya çıkaran ve şirketlerin büyük ölçekte satış yapmasını amaçlayan bir yöntemdir.
- Örneğin bir marketten müşterilerin süt ve peynir satın alımlarının %70'inde bu ürünler ile birlikte yoğurtta satın alınmıştır. Pazar sepeti analizi bu ilişkiyi ortaya çıkarır ve satıcılar buna göre bir satış politikası izlerler.
- Tüketicilerin satın alma davranışları, alışveriş alışkanlıklarının ve birlikte aldıkları ürünlerin incelenmesiyle elde edilir.
- İşletmeler daha fazla ürün satabilmek için müşteri alışkanlıkları bilmek ve ortaya çıkarmak istemektedir.
- Amerika'da e-ticaret devi **Amazon**'un gelirlerinin ciddi bir kısmını (yaklaşık %75'i) bu şekilde elde ederken, **Netflix** ise yaklaşık %35'ini elde eder.

- Bunlara ek, **Spotify** kullanıcılarına tavsiye olarak önereceği müzikleri kullanıcıların alışkanlıklarını analiz edip ürün ve kullanıcı bazlı benzerliklerden yararlanarak kullanıcılara sunar.
- Tüm bunların yanında, "sıklıkla birlikte alınan ürünler" ve "bu ürünü alanlar bunu da aldı" gibi başlıklar da pazar sepeti analizinin bir sonucudur.

#### Frequently Bought Together

Price for all three: **\$74.20**

[Add all three to Cart](#) [Add all three to Wish List](#)

[Show availability and shipping details](#)

✓ **This item:** Beginning Ruby: From Novice to Professional (Expert's Voice in Open Source) by Peter Cooper Paperback **\$27.78**

✓ **Learn to Program, Second Edition (The Facets of Ruby Series)** by Chris Pine Paperback **\$16.94**

✓ **Ruby on Rails Tutorial: Learn Web Development with Rails (2nd Edition) (Addison-Wesley Professional Ruby ...)** by Michael Hartl Paperback **\$29.48**

#### Customers Who Bought This Item Also Bought

Learn to Program, Second Edition (The Facets of Ruby Series) by Chris Pine  
★★★★★ 42  
Paperback  
**\$16.94** ✓Prime

The Well-Grounded Rubyist by David A. Black  
★★★★★ 39  
Paperback  
**\$32.49** ✓Prime

Ruby on Rails Tutorial: Learn Web Development with Rails (2nd Edition) (Addison-Wesley Professional Ruby Series) by Michael Hartl  
★★★★★ 70  
Paperback  
**\$29.48** ✓Prime

The Ruby Programming Language by David Flanagan  
★★★★★ 74  
Paperback  
**\$26.35** ✓Prime

The Well-Grounded Rubyist by David A. Black  
★★★★★ 19  
**#1 Best Seller** in Ruby Programming Computer  
Paperback  
**\$29.67** ✓Prime

### MentholMusic

|   | SONG           | ARTIST                  | ALBUM  |
|---|----------------|-------------------------|--------|
| + | Winter         | North Elements, AaA     | Winter |
| + | Fading Through | Wafia, Vancouver Sle... | XXIX   |

### Recommended Songs

Based on the songs in this playlist

|   |                  |                 |               |
|---|------------------|-----------------|---------------|
| + | The Breach       | Dustin Tebbutt  | The Breach EP |
| + | Portrait 19      | The Paper Kites | States        |
| + | This Is the Walk | David O'Dowda   | Look Up       |
| + | Settle           | Vera Blue       | Settle        |
| + | For My Help      | Hayden Calnin   | City          |
| + | From Gold        | Novo Amor       | From Gold     |

# Pazar Sepeti Uygulamaları

- **Çapraz Satış Yapılması(Cross-Sell):** Bir ürünü satarken, bu ürünü tamamlayacak bir yan ürünü, ürünle beraber tüketiciye sunarak, tüketicinin birden fazla ürünü satın almaya yönlendirecek satış yöntemidir. Tüketicinin, “Sepet” sayfası üzerinde de kullanıcının sepete eklediği ürünle ilgili olabilecek alternatif ürünleri listelenmesi bir çapraz satış örneğidir.

## Sepetim (1 ürün)

Satıcı: T-Gex >



Asus ROG Strix G513IC-HN037 AMD Ryzen 7 4800H 16GB 512GB SSD

RTX3050 Freedos 15.6" 144 Hz FHD Taşınabilir Bilgisayar • Gri

En geç 12 Aralık Pazartesi günü kargoda

22.699,00 TL



1



## Sıklıkla birlikte alınanlar



Gamingpadci Purple Wave  
Desenli 90X40 cm...

★★★★★ 63

89,10 TL



Rampage AD-RX44 X-Gale  
Siyah 6 Fanlı Notebook...

★★★★★ 240

639,00 TL



Katsuta T4-1618 Mortimus  
Makro Tuşlu RGB Işıklı...

★★★★★ 217

219,99 TL



Pusat Hafif Oyuncu Kulaklığı -  
Siyah

★★★★★ 207

499,00 TL

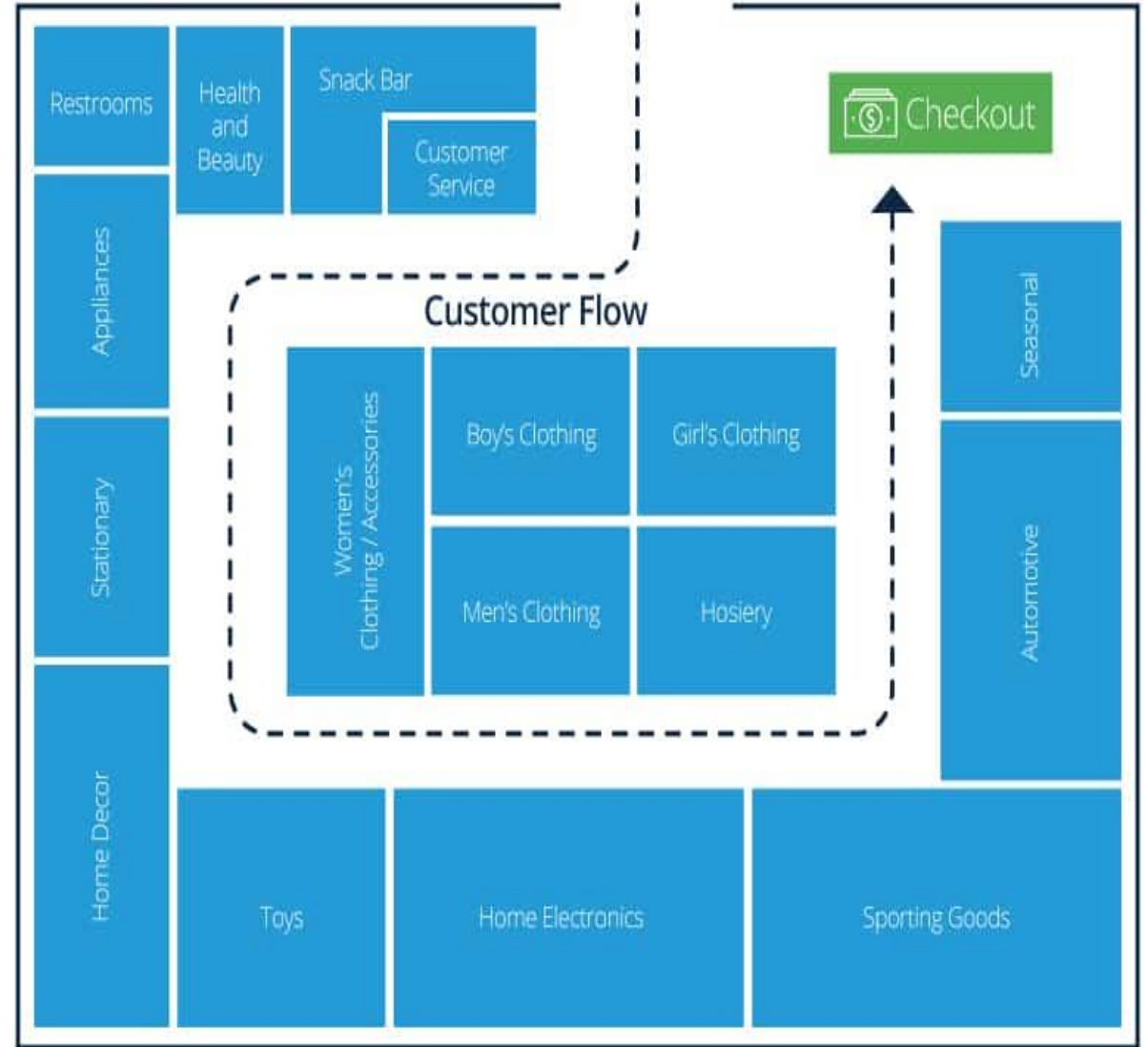


Mack MCC-007 17.3" Offis  
USB Girişli Notebook Sirt..

★★★★★ 133

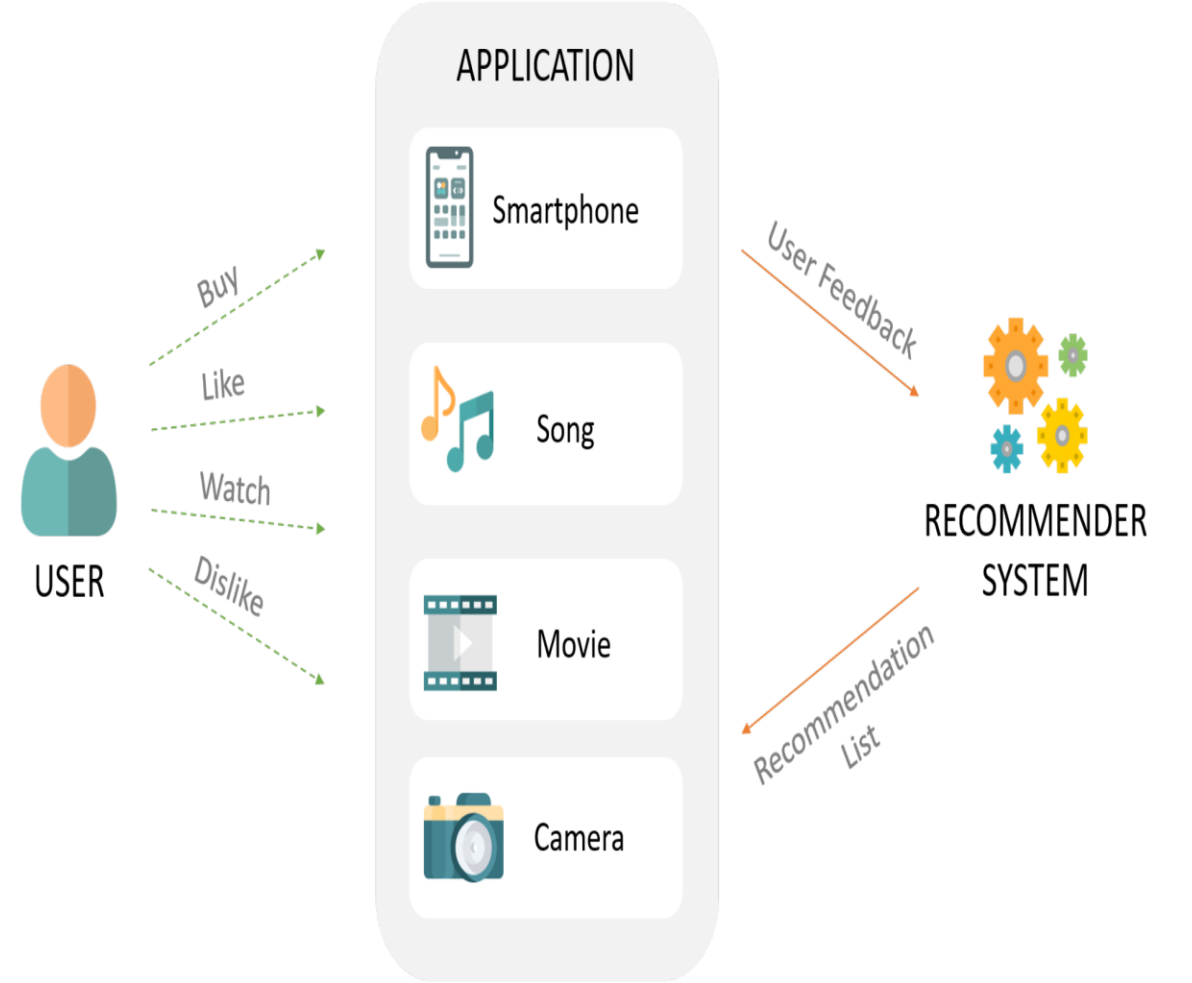
374,90 TL

- **Mağaza Düzeni(Store Layout):** Satışları ve geliri artırmak için mağazadaki ürünler pazar sepeti analizine göre düzenlenebilir.
- Pazar sepeti analizine göre ürünler müşterinin dikkatini çekecek ve diğer ürünü satın alma kararına yardımcı olacak şekilde sıralanabilir veya yerleştirilebilir.
- Pazar sepeti analizi, satışların artmasıyla sonuçlanan ürünleri ve kombinasyonları düzenlemek için bir rehber görevi görür.





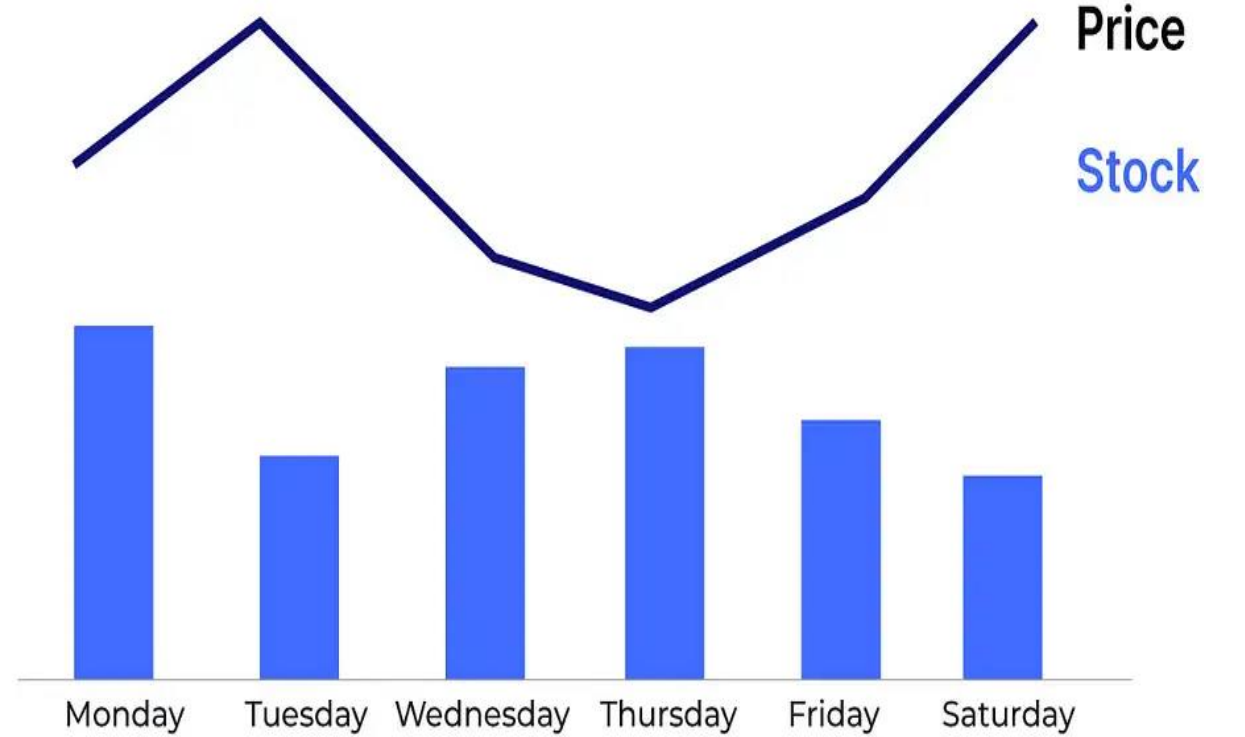
- **Tavsiye Motorları(Recommendation Engines)**
- Pazar sepeti analizi, tavsiye motorlarını güçlendiren temel ilkedir.
- Tavsiye sistemleri, kullanıcıların deneyimlerini, davranışlarını, tercihlerini ve ilgi alanlarını kullanarak tüm dijital dünyayı keşfetmelerinin birincil yolu haline geliyor. Bilgi yoğunluğunun ve aşırı ürün alışverişinin olduğu bir dünyada, bir tavsiye motoru, şirketlere tüketicilere kişiselleştirilmiş bilgi ve çözümler sunmanın verimli bir yolunu sunar.





- **Fiyatlandırma Faktörü ve Satış Stokları(Pricing Factor and Sales Inventory):**

- Stok bazlı fiyatlandırma, fiyatların stok seviyelerine veya tahminlere göre ayarlandığı bir dinamik fiyatlandırma şeklidir. Bir ürünün fiyatını mevsimsellik, haftanın günü, rakip şirket fiyatları vb. dahil olmak üzere birçok faktör belirlerken, belirli bir ürünün elinizde bulunan gerçek miktarı, fiyatları hesaplarken son derece önemlidir.
- Örnek olarak belirli yaş grubundaki annelerin süt tozu aldıklarında, aynı zamanda bebek bezi satın aldıklarını varsayarsak bu, perakendecilerin fiyatları belirlemesine ve ayrıca satış hacmini korumasına yardımcı olur.



# Apriori Algoritması kullanarak Pazar Sepeti Analizi

- Apriori algoritması, nesneler arasındaki pazar sepeti analizini hesaplamak için kullanılan algoritmayı ifade eder.
- Apriori algoritmasını hesaplamak için üç kavram vardır:

**1- Support:** X ve Y ürünlerinin birlikte satın alınması olasılığını belirtir.

$$\text{Support}(X, Y) = \text{frequency}(X, Y) / N$$

**2 – Confidence:** X ürününü alanların Y ürününü alma olasılığını belirtir.

$$\text{Confidence}(X, Y) = \text{frequency}(X, Y) / \text{frequency}(X)$$

**3 – Lift:** X'in satın alınması durumunda Y'nin satın alınmasının kaç kat artacağını belirtir.

$$\text{Lift}(X, Y) = \text{Confidence}(X, Y) / (\text{Support}(X))$$

| İşlem         | Ürünler                  |
|---------------|--------------------------|
| Transaction 1 | Ekmek, Süt, Kola, Peynir |
| Transaction 2 | Ekmek, Süt, Kola         |
| Transaction 3 | Ekmek, Süt               |
| Transaction 4 | Ekmek, Soğan             |
| Transaction 5 | Un, Süt, Kola, Peynir    |
| Transaction 6 | Un, Süt, Kola            |
| Transaction 7 | Un, Süt                  |
| Transaction 8 | Un, Soğan                |

$$\text{Support}(\text{Ekmek}) = \frac{4}{8} \quad \text{Support}(\text{Süt}) = \frac{6}{8}$$

$$\text{Support}(\text{Ekmek}, \text{Süt}) = \frac{3}{8}$$

$$\text{Confidence}(\text{Ekmek}, \text{Süt}) = \frac{\text{Freq}(\text{Ekmek}, \text{Süt})}{\text{Freq}(\text{Ekmek})} = \frac{3}{4}$$

$$\text{Lift}(\text{Ekmek}, \text{Süt}) = \frac{\text{Support}(\text{Ekmek}, \text{Süt})}{\text{Support}(\text{Ekmek}) * \text{Support}(\text{Süt})} = \frac{\frac{3}{8}}{\frac{4}{8} * \frac{6}{8}} = 1$$

lift < 1 : ürünlerin satışı birbirini negatif olarak etkiler.

lift = 1 : ürünler birbirine bağımlı değildir.

lift > 1 : ürünler arasında pozitif ilişki vardır.

\*\*\* Yukarıdaki örnekte lift değeri 1 olduğu için ekmek ve süt ürünleri arasında bir ilişki yoktur.

```
[] resultsinDataFrame
```



|   | Left Hand Side       | Right Hand Side | Support  | Confidence | Lift     |
|---|----------------------|-----------------|----------|------------|----------|
| 0 | light cream          | chicken         | 0.004533 | 0.290598   | 4.843951 |
| 1 | mushroom cream sauce | escalope        | 0.005733 | 0.300699   | 3.790833 |
| 2 | pasta                | escalope        | 0.005866 | 0.372881   | 4.700812 |
| 3 | fromage blanc        | honey           | 0.003333 | 0.245098   | 5.164271 |
| 4 | herb & pepper        | ground beef     | 0.015998 | 0.323450   | 3.291994 |
| 5 | tomato sauce         | ground beef     | 0.005333 | 0.377358   | 3.840659 |
| 6 | light cream          | olive oil       | 0.003200 | 0.205128   | 3.114710 |
| 7 | whole wheat pasta    | olive oil       | 0.007999 | 0.271493   | 4.122410 |
| 8 | pasta                | shrimp          | 0.005066 | 0.322034   | 4.506672 |

- \*Fromage blanc >> peynir türü, honey >> bal
- $\text{Support}(\text{fromage blanc}, \text{honey}) = 0.0033$   $\text{Confidence}(\text{fromage blanc}, \text{honey}) = 0.245$ ,  $\text{Lift}(\text{fromage blanc}, \text{honey}) = 5.164$
- Yapılan 1000 alışverişin yaklaşık %3'ünde peynir ve bal birlikte alınmıştır. Peynir alanların yaklaşık %25' i bal da almıştır. Peynirin satılması balın satılmasını 5 kat daha artırmıştır.

# KAYNAKÇA

- <https://www.neenopal.com/MarketBasketAnalysis.html>
- <https://www.slideshare.net/uslumetin/birliktelik-kurallar-kullanilarak-pazar-sepeti-analizi-market-basket-analysis-using-association-rules>
- <https://djinit-ai.github.io/2020/09/22/apriori-algorithm.html>
- <https://medium.com/@eminenurnacar/veri-madencili%C4%9Finde-pazar-sepeti-analizi-de44962f2086>
- [https://www.google.com/url?sa=i&url=https%3A%2F%2Flearnings.com%2Fblog%2Finventory-based-pricing%2F&psig=AOvVaw2\\_Dfl4R5AuyLPJzYu0sPch&ust=1670836738093000&source=images&cd=vfe&ved=0CBAQjRxqFwoTCKjv9oWe8fsCFQAAAAAdAAAAABAD](https://www.google.com/url?sa=i&url=https%3A%2F%2Flearnings.com%2Fblog%2Finventory-based-pricing%2F&psig=AOvVaw2_Dfl4R5AuyLPJzYu0sPch&ust=1670836738093000&source=images&cd=vfe&ved=0CBAQjRxqFwoTCKjv9oWe8fsCFQAAAAAdAAAAABAD)
- <https://www.veribilimiokulu.com/associationrulesanalysis/>
- [https://cdn-gcp.new.marutitech.com/e8910875\\_cacf6c36\\_03dbeffa\\_frequently\\_bought\\_together\\_min\\_705x423\\_c403c2789e.png](https://cdn-gcp.new.marutitech.com/e8910875_cacf6c36_03dbeffa_frequently_bought_together_min_705x423_c403c2789e.png)
- <https://medium.com/machine-learning-t%C3%BCrkiye/ad%C4%B1m-ad%C4%B1m-makine-%C3%B6%C4%9Frenmesi-b%C3%B6l%C3%Bcm-3-denetimsiz-%C3%B6%C4%9Frenme-nedir-f89oada49a40>
- <https://yavuz.github.io/denetimsiz-ogrenme/>
- [https://www.youtube.com/watch?v=CMbLMiJAx\\_g](https://www.youtube.com/watch?v=CMbLMiJAx_g)
- <https://samed-harman.medium.com/makine-%C3%B6%C4%9Frenmesi-clustering-k%C3%BCmeleme-teknikleri-bd1b59a0a177>
- <https://www.btkakademi.gov.tr/portal/course/player/deliver/python-ile-makine-ogrenmesi-11800>

# KAYNAKÇA

- <https://samed-harman.medium.com/makine-%C3%B6%C4%9Frenmesi-clustering-k%C3%BCmeleme-teknikleri-bd1b59a0a177>
- <https://medium.com/@gulcanogundur/pca-principal-component-analysis-temel-bile%C5%9Fenler-analizi-bf9098751c62>
- <https://medium.com/data-science-tr/makine-%C3%B6%C4%9Frenmesi-dersleri-boyut-azaltma-pca-5ae9e902ef92>
- <https://samed-harman.medium.com/makine-%C3%B6%C4%9Frenmesi-clustering-k%C3%BCmeleme-teknikleri-bd1b59a0a177>
- [https://erdincuzun.com/makine\\_ogrenmesi/hiyerarsik-kumeleme-hierarchical-clustering-odev-benzerlikleri-uzerinden-kopya-gruplarini-bulma/](https://erdincuzun.com/makine_ogrenmesi/hiyerarsik-kumeleme-hierarchical-clustering-odev-benzerlikleri-uzerinden-kopya-gruplarini-bulma/)
- <https://www.btkakademi.gov.tr/portal/course/player/deliver/python-ile-makine-ogrenmesi-11800>
- [https://www.google.com/imgres?imgurl=https%3A%2F%2Fwww.veribilimiokulu.com%2Fwp-content%2Fuploads%2F2019%2F10%2FKumeleme\\_Notlari\\_4\\_K-Ortalamalar\\_Python\\_Uygulama\\_Kumeler\\_grafik\\_gosterim.png&imgrefurl=https%3A%2F%2Fwww.veribilimiokulu.com%2FKumeleme-notlari-4-k-ortalamalar-python-uygulama%2F&tbnid=IQsiD7CoJHQi2M&vet=12ahUKEWjl2PLH\\_\\_b7AhXdIMUKHU-XApoQMyhgegUIARCxAQ..i&docid=oP8fBezONugU3M&w=398&h=281&q=k%2omeans%2ok%C3%BCmeleme%2oalgoritmas%C4%B1%2ok%2omean&hl=tr&ved=2ahUKEWjl2PLH\\_\\_b7AhXdIMUKHU-XApoQMyhgegUIARCxAQ](https://www.google.com/imgres?imgurl=https%3A%2F%2Fwww.veribilimiokulu.com%2Fwp-content%2Fuploads%2F2019%2F10%2FKumeleme_Notlari_4_K-Ortalamalar_Python_Uygulama_Kumeler_grafik_gosterim.png&imgrefurl=https%3A%2F%2Fwww.veribilimiokulu.com%2FKumeleme-notlari-4-k-ortalamalar-python-uygulama%2F&tbnid=IQsiD7CoJHQi2M&vet=12ahUKEWjl2PLH__b7AhXdIMUKHU-XApoQMyhgegUIARCxAQ..i&docid=oP8fBezONugU3M&w=398&h=281&q=k%2omeans%2ok%C3%BCmeleme%2oalgoritmas%C4%B1%2ok%2omean&hl=tr&ved=2ahUKEWjl2PLH__b7AhXdIMUKHU-XApoQMyhgegUIARCxAQ)
- <https://medium.com/machine-learning-türkiye/adım-adım-makine-öğrenmesi-bölüm-3-denetimsiz-öğrenme-nedir-f89oada49a40>
- <https://www.ibm.com/cloud/learn/unsupervised-learning#:~:text=Unsupervised%20learning%2C%2oalso%2oknown%2oas,the%2oneed%2ofor%2ohuman%2ointervention.>
- <https://www.expert.ai/blog/machine-learning-definition/>