

Association rule mining

Association rule learning/mining = birliktelik kural çıkarımı

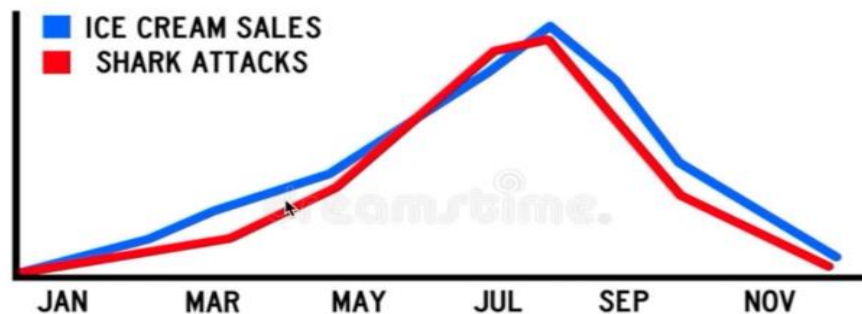
ARM / ARL



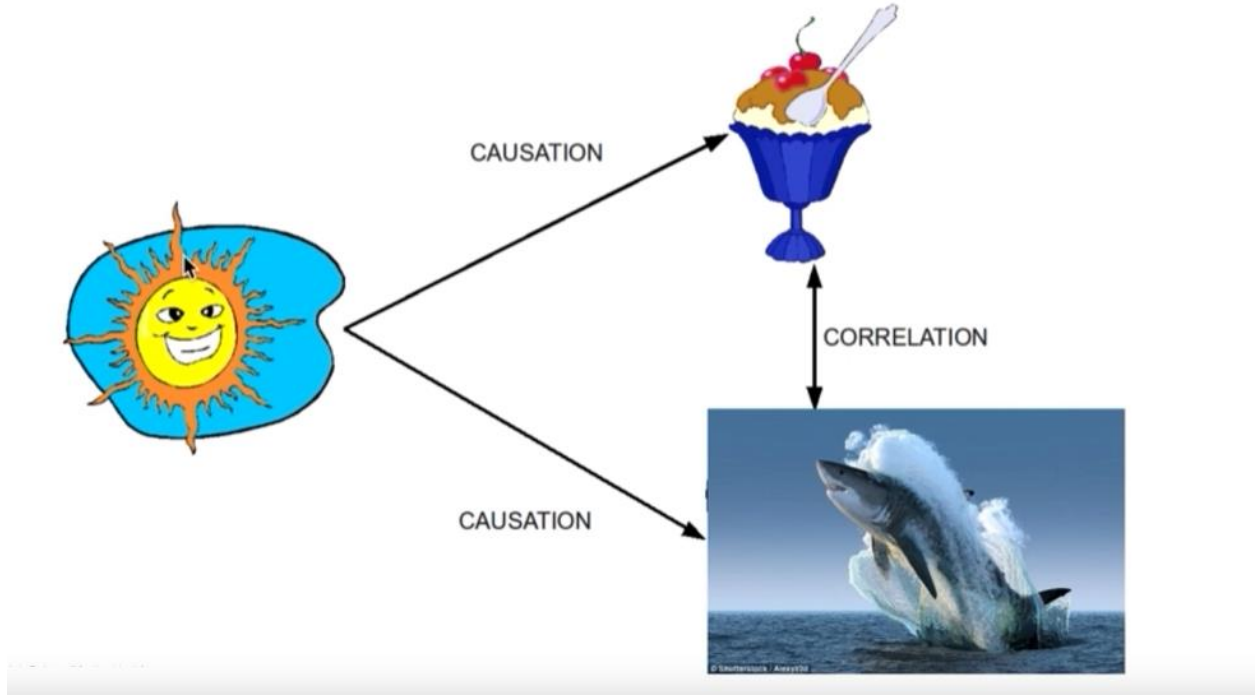
Bunu alanlar, bunu da aldı
Bunu izleyenler bunu da izledi
Bunu ...'lar bunu da ...

CORRELATION VS CAUSALITY

CORRELATION IS NOT CAUSATION!



Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other (they are caused by good weather, with lots of people at the beach, both eating ice cream and having a swim in the sea)



Örnek

- Veri Tabanı

Transaction	İçerik
T1	a,b,c
T2	a,c
T3	a,e,c
T4	f,a
T5	f,e

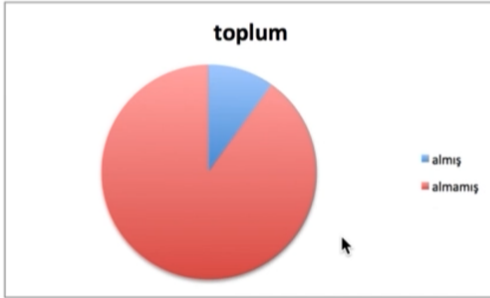
- Kurallar

$X \Rightarrow Y$	S	A
$a \rightarrow c$	%60	%75
$c \rightarrow a$	%60	%100
$e \rightarrow a$	%20	%50
$c \rightarrow b$	%20	%33.3
$b \rightarrow c$	%20	%100
$b \rightarrow f$	%0	%0

Bir veri tabanı üzerinden anlamsal, ilişkisel değerler çıkarmaya çalışıyor.

Destek (Support)

Yüz kişiye sorduk



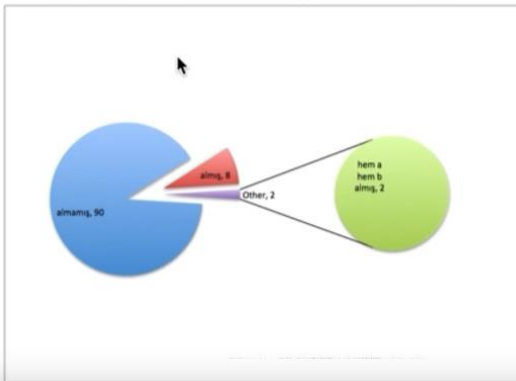
$$\text{Support}(a) = \frac{\text{a varlığını içeren eylemler}}{\text{Toplam Eylem Sayısı}}$$

A varlığı = X filmini izlemek olsun [mavi kısım]

$$\text{Support}(a) = 10/100 = \%10$$

Yüz kişiye sorduk

a ürününü alan 100 kişiye sorduk



$$\text{Support}(a) = \frac{\text{a varlığını içeren eylemler}}{\text{Toplam Eylem Sayısı}}$$

$$\text{Confidence}(a \rightarrow b) = \frac{\text{a ve b varlığını içeren eylemler}}{\text{a varlığını içeren eylemler}}$$

B varlığı = Y Filmini izlemek olsun [mor]

A ve B varlığını içeren eylemler[a izleyenler b de izlemiştir] = 2

$$\text{Confidence}(a \rightarrow b) = 2/10 = \%20$$

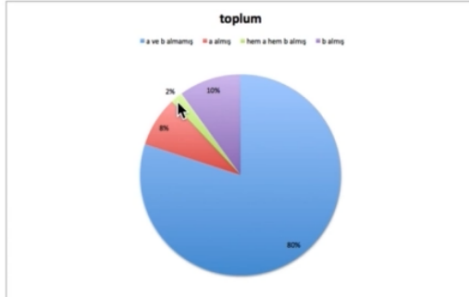
Destek (Support)

Yüz kişiye sorduk

$$\text{Support}(a) = \frac{\text{a varlığını içeren eylemler}}{\text{Toplam Eylem Sayısı}}$$

a ürününü an 100 kişiye sorduk

B ürününü amak ne kadar artış sağlar?



$$\text{Confidence}(a \rightarrow b) = \frac{\text{a ve b varlığını içeren eylemler}}{\text{a varlığını içeren eylemler}}$$

$$\text{Lift}(a \rightarrow b) = \frac{\text{Confidence}(a \rightarrow b)}{\text{support}(b)}$$

B ürününü alanlar = 12

$\text{Confidence}(a \rightarrow b) = 2/10 = \%20$

$\text{Support}(b) = 12/100 = \%12$

Ancak;

$\text{Lift}(a \rightarrow b) = 20/12 = 1.67$

Liftin 1'den büyük olması a'nın alınması b ürünün alınmasını olumlu yönde etkiliyor

1'in altındaysa olumsuz etkiliyor demektir.

Lift = [a ürününü alıp b ürününü de alanlar %20 (10 kişinin 2 si b ürününü de almış), yani b ürününü alma oranı %12'den %20ye çıktı a->b yapınca]

APriori Algoritması

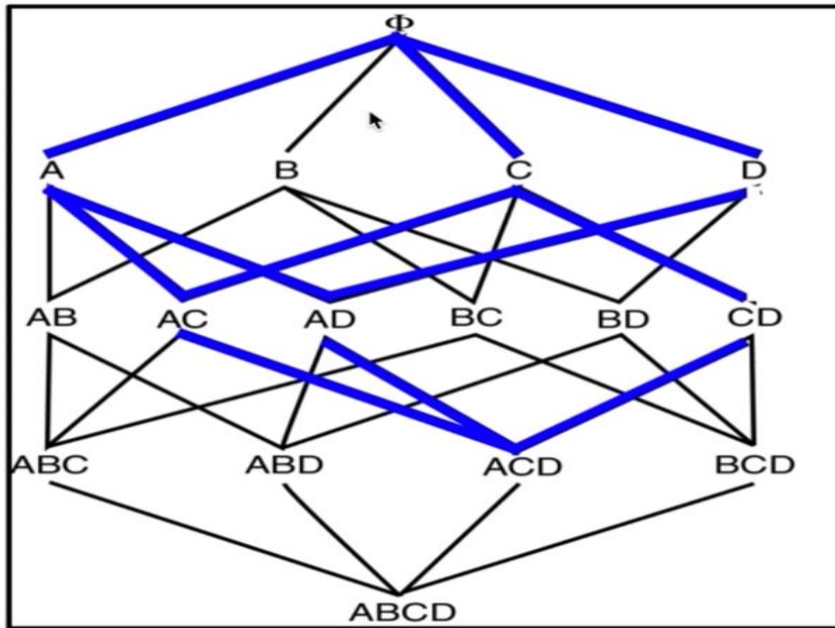
Apriori algoritması, birliktelik kuralı çıkarımı için yaygın olarak kullanılan bir veri madenciliği algoritmasıdır. Bu algoritma, bir veri kümesindeki sık olarak görülen öğe kümelerini (itemsets) bulmak için kullanılır. Özellikle, belirli bir destek (support) eşiği üzerindeki sık öğe kümeleri bulunur.

Apriori algoritması genellikle üç adımdan oluşur:

1. **Küçük öğe kümelerini bulma (itemset generation):** Başlangıçta, tek öğelerin destek değerlerini hesaplayarak sık öğelerin bir listesini oluşturur.
2. **Birleştirme (joining):** Sık öğelerden oluşturulan aday öğe kümeleri, daha büyük öğe kümeleri oluşturmak için birleştirilir.
3. **Eleme (pruning):** Oluşturulan aday öğe kümeleri, belirlenen destek eşiğini karşılamayanları eleme adımında filtrelendir. Bu şekilde, yalnızca sık öğe kümeleri alınır.

Bu adımlar, veri kümesindeki sık öğe kümelerini bulmak için tekrarlanır. Sonuç olarak, Apriori algoritması sık olarak birlikte görülen öğeleri (örneğin, bir market alışverişinde birlikte satın alınan ürünleri) belirlemek için kullanılır. Bu bilgi, örneğin pazarlama stratejilerini belirlemek veya ürün yerleşimini optimize etmek gibi alanlarda kullanılabilir.

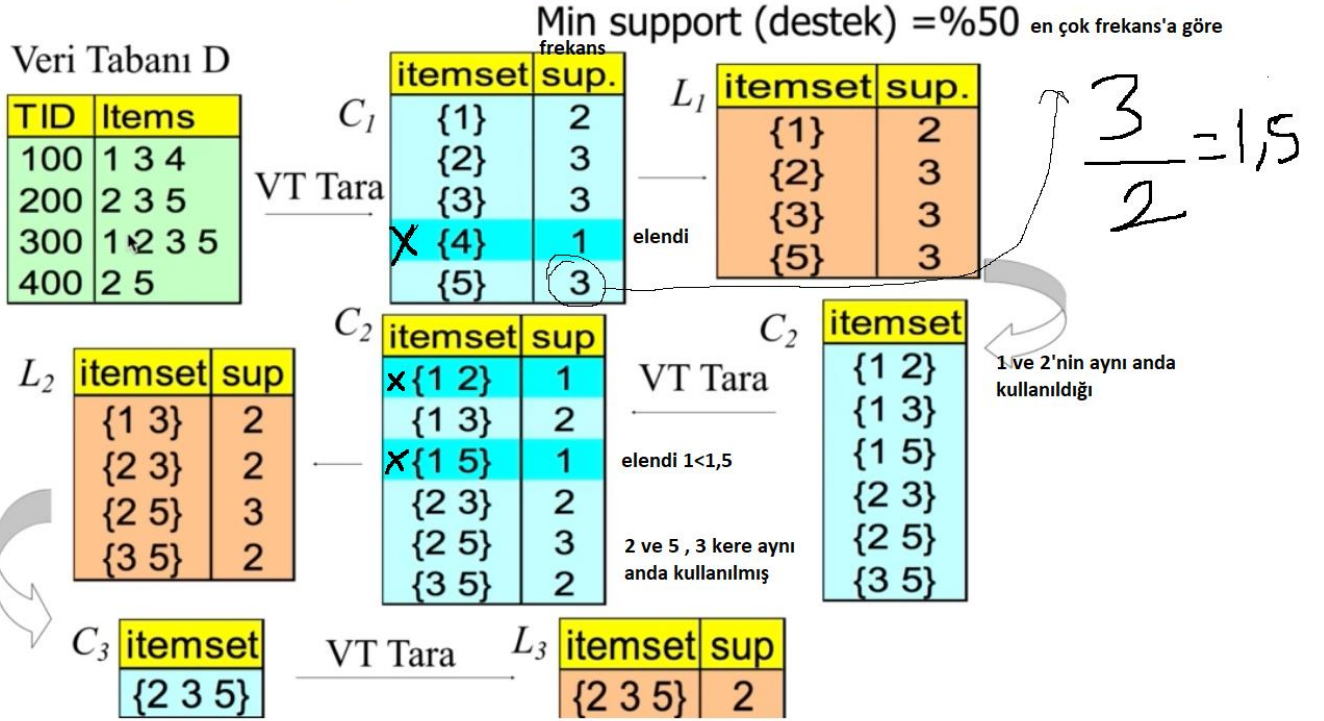
Olay Sıklığı



4 farklı ürünün(olayın) tek tek support değerlerine bakıyor ona göre eleme yapıyor.

Support değerleri belirlenen eşik üstündeyse ikili durumların değerleri inceleniyor(a->b gibi)

Apriori Algoritması — Örnek



ARM / ARL



Nerelerde kullanılır?

Complex Event Processing
Kampanya
Davranış Tahmini
Yönlendirilmiş ARM
Zaman Serisi Analizi

Yönlendirilmiş ARM= bir ürün alırken a->b a ile b nin alınması b->a şeklinde gösterilebilir yönsüzdür fakat a'dan sonra b (a->b) eyleminin yapılması b->a olarak gösterilemez yönlü bir işlemdir

ECLAT Algoritması

Equivalence Class Transformation

Apriori : Breadth First Search

Eclat : Depth First Search

Eclat Algoritması — Örnek

Min support (destek) = %50

Veri Tabanı D

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

VT Tara

1	2	3	4	5
100	200	100	100	200
300	400	200		300
		300		400

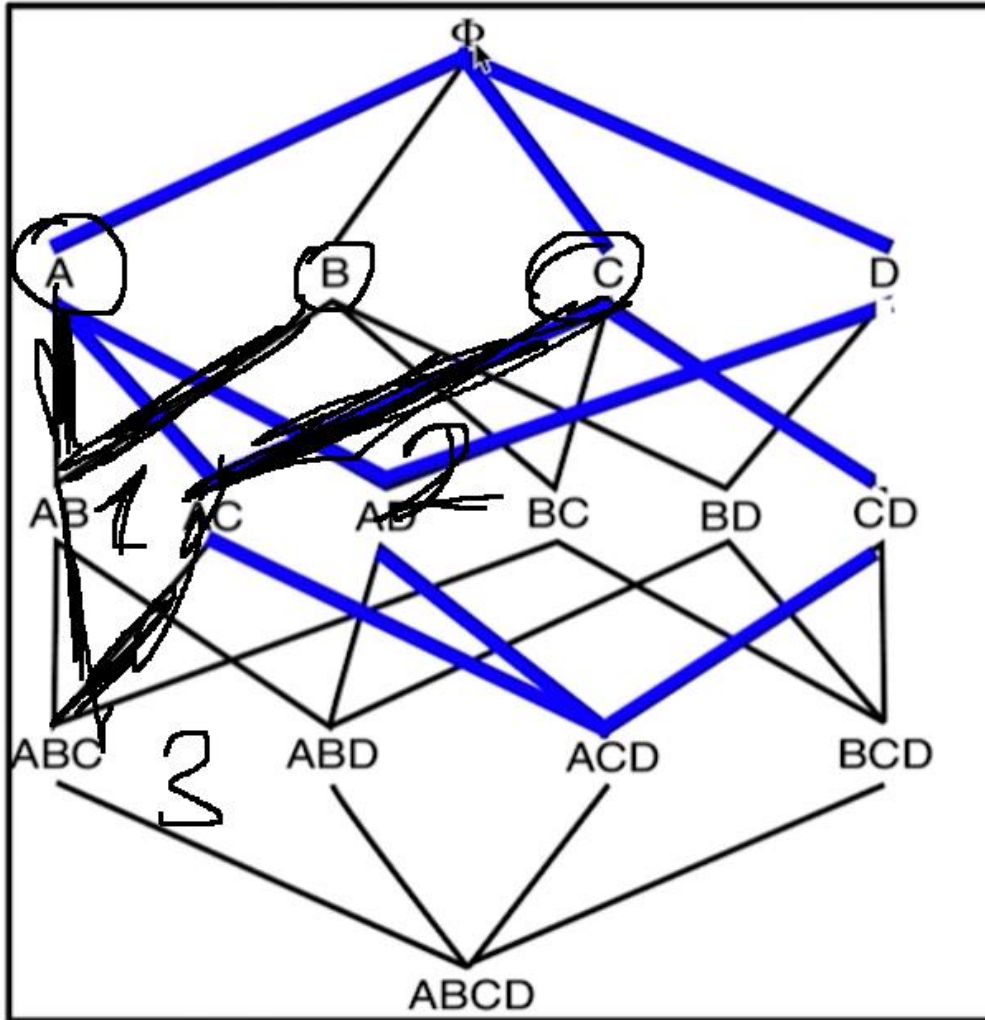
12	13	15	23	25	35
	100	300	200	200	200
	300			400	300

135	235
300	200

eclat derin öncelikli : DFS

Olay Sıklığı

ÖNCE DERİNE



APriori ise yayılım(genişlik) öncelikli : BFS

Önce A-B-C-D support değerleri hesaplanacak sonra bir aşağıya

Fakat eclat algoritmasında

A-B hesaplandıktan sonra AB c hesaplandıktan sonra abc şeklinde derin

Eclat daha hızlı.