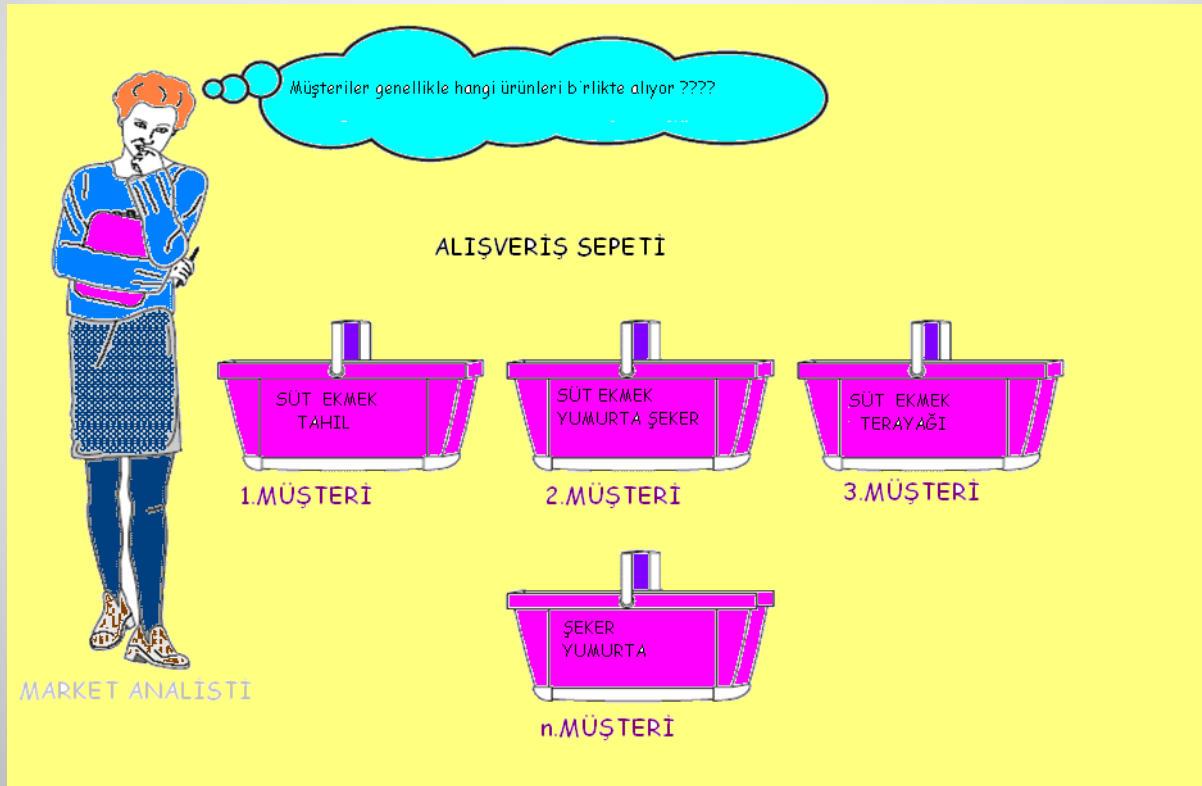


Veri Madenciliđi

BİRLİKTELİK KURALLARI VE İLİŞKİ ANALİZİ



Birliktelik kuralları, aynı işlem içinde çoğunlukla beraber görülen nesneleri içeren kurallardır. Birliktelik kurallarının kullanıldığı en tipik örnek market sepeti uygulamasıdır. Bu işlem, müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikleri bularak **müşterilerin satın alma alışkanlıklarını çözümler**.



$$\text{destek}(X \rightarrow Y) = \frac{\text{sayı}(X, Y)}{n}$$

$$\text{güven}(X \rightarrow Y) = \frac{\text{sayı}(X, Y)}{\text{sayı}(X)}$$

•Örneğin 25 tane müşterinin bir defada aldığı ürün bilgilerinden yola çıkarak birliktelik kuralı şu şekilde bulunmuş olsun:

$$\text{güven}(\text{Pantolon}, \text{Kazak} \rightarrow \text{Çorap})$$

•Burada $X = \{\text{Pantolon}, \text{Kazak}\}$ ve $Y = \{\text{Çorap}\}$

değerleri için pantolon ve kazak alan müşterilerin bunların yanında çorap da sayın alama olasılığını ifade eder. Müşterinin bu 3 ürünü birlikte satın alma sayısı 7 ve müşteri sayısı 25 ise belirttiğimiz bu kuralın destek ölçütü şöyle olacaktır:

$$\text{destek}(\text{Pantolon}, \text{Kazak} \rightarrow \text{Çorap}) = \frac{\text{sayı}(\text{Pantolon}, \text{Kazak}, \text{Çorap})}{\text{musterisayisi}} = \frac{7}{25} = 0,28$$

•Eğer pantolon ve kazak alanların sayısının 9 ise Güven ölçütü

$$\text{güven}(\text{Pantolon}, \text{Kazak} \rightarrow \text{Çorap}) = \frac{\text{sayı}(\text{Pantolon}, \text{Kazak}, \text{Çorap})}{\text{sayı}(\text{Pantolon}, \text{Kazak})} = \frac{7}{9} = 0,77$$

Apriori Algoritması

Bu algoritma birliktelik kurallarının oluşturulmasında yararlanılan ve yaygın olarak kullanılan bir algoritmadır. Algoritma aşağıdaki adımlardan oluşmaktadır:

- *Öncelikle destek ve güven ölçülerini karşılaştırmak için eşik değerleri belirlenir.*
- *Her bir ürün için destek sayılan hesaplanır. Eşik değeri ile karşılaştırılan destek değerlerinin içinden eşik değerinden düşük olanlar çıkarılır.*
- *Kalan ürünler ikiyeşerli gruplanarak, grup destek sayılan hesaplanır. Tekrar eşik değerleri ile karşılaştırılan destek değerlerinden eşik değerinin altında kalanlar iptal edilir.*
- *Daha sonra üçerli, dörderli, beşerli, vb. biçimde gruplar için aynı karşılaştırma ve eleme işlemi devam ettirilir. Eşik değerlere uygun olduğu sürece işlemler sürecektir.*
- *Belirlenen ürün grubunun destek ölçülerine bakarak birliktelik kuralları türetilir ve bu kurallarının her biri için güven ölçüleri belirlenir.*

Apriori Algoritması(Örnek Çalışma)

- Öncelikle destek ve güven ölçülerini karşılaştırmak için eşik değerleri belirlenir.

Müşteri Numarası	Aldığı Çikolata Markası
MSN0101	ERDEM, ŞOKOKO, ÇİTİT, DERYA
MSN0102	ENFES, DERYA, MAZLUM, ŞOKOKO
MSN0103	ERDEM, DERYA, ENFES
MSN0104	ŞOKOKO, ÇİTİT, DERYA, POTPORİ,
MSN0105	ERDEM, DERYA, MAZLUM, ÇİTİT
MSN0106	MİS, DERYA, ÇİTİT
MSN0107	ERDEM, ZARİF, DERYA
MSN0108	MAZLUM, YURDUM, ÇİTİT
MSN0109	YURDUM, ŞOKOKO, ÇİTİT
MSN0110	ÇİTİT, MİS, ERDEM, MAZLUM, DERYA

$$\text{Destek}_{\text{eşik}} = \%30$$

$$\text{Güven}_{\text{eşik}} = \%80$$

Burada eşik destek sayısı

$$0,30 * 10 = 3$$

Apriori Algoritması(Örnek Çalışma)

- Her bir ürün için destek sayıları hesaplanır.Eşik değeri ile karşılaştırılan destek değerlerinin içinden eşik değerinden düşük olanlar çıkarılır.

<u>Cikolata Markası</u>	<u>Destek Değeri</u>
ERDEM	5
ŞOKOKO	4
ÇİTPİT	7
MAZLUM	4
ENFES	2
DERYA	8
POTPORI	1
MİS	2
YURDUM	2
ZARIF	1

<u>Cikolata Markası</u>	<u>Destek Değeri</u>
ERDEM	5
ŞOKOKO	4
ÇİTPİT	7
MAZLUM	4
DERYA	8

Apriori Algoritması(Örnek Çalışma)

Çikolata Markası	Destek Değeri
ERDEM_SOKOKO	1
ERDEM_QITPIT	3
ERDEM_MAZLUM	2
ERDEM_DERYA	5
SOKOKO_ERDEM	1
SOKOKO_QITPIT	3
SOKOKO_MAZLUM	2
SOKOKO_DERYA	3
QITPIT_ERDEM	3
QITPIT_SOKOKO	3
QITPIT_MAZLUM	3
QITPIT_DERYA	5
MAZLUM_ERDEM	2
MAZLUM_SOKOKO	2
MAZLUM_QITPIT	3
MAZLUM_DERYA	3
DERYA_ERDEM	5
DERYA_QITPIT	5
DERYA_MAZLUM	3
DERYA_SOKOKO	3

- Kalan ürünler ikışerli gruplanarak, grup destek sayıları hesaplanır. Tekrar eşik değeri ile karşılaştırılan destek değerlerinden eşik değeri altında kalanlar iptal edilir.

Çikolata Markası	Destek Değeri
ERDEM_QITPIT	3
ERDEM_DERYA	5
SOKOKO_QITPIT	3
SOKOKO_DERYA	3
QITPIT_MAZLUM	3
QITPIT_DERYA	5
MAZLUM_DERYA	3

Apriori Algoritması(Örnek Çalışma)

- Daha sonra üçerli,dörderli,beşerli, vb. biçimde gruplar için aynı karşılaştırma ve eleme işlemi devam ettirilir. Eşik değerlere uygun olduğu sürece işlemler sürecektir.

Çikolata Markası	Destek Değeri
ERDEM, ÇİTİT, ŞOKOKO	1
ERDEM, ÇİTİT, DERYA	3
ERDEM, DERYA, ŞOKOKO	2
ERDEM, DERYA, ÇİTİT	3
ERDEM, DERYA, ENFES	1
ERDEM, DERYA, MAZLUM	3
ERDEM, DERYA, MİS	1
ŞOKOKO, ÇİTİT, ERDEM	1
ŞOKOKO, ÇİTİT, DERYA	2
ŞOKOKO, ÇİTİT, POTPORİ	1
ŞOKOKO, ÇİTİT, YURDUM	1
ŞOKOKO, DERYA, ERDEM	1
ŞOKOKO, DERYA, ÇİTİT	2
ŞOKOKO, DERYA, MAZLUM	1

ŞOKOKO, DERYA, ENFES	1
ŞOKOKO, DERYA, POTPORİ	1
ÇİTİT, MAZLUM, DERYA	2
ÇİTİT, MAZLUM, ERDEM	2
ÇİTİT, MAZLUM, YURDUM	1
ÇİTİT, MAZLUM, MİS	1
ÇİTİT, DERYA, ŞOKOKO	2
ÇİTİT, DERYA, ERDEM	3
ÇİTİT, DERYA, POTPORİ	1
ÇİTİT, DERYA, MAZLUM	2
ÇİTİT, DERYA, MİS	2

Çikolata Markası	Destek Değeri
ERDEM, ÇİTİT, DERYA	3
ERDEM, DERYA, MAZLUM	3

Apriori Algoritması(Örnek Çalışma)

- Belirlenen ürün grubunun destek ölçülerine bakarak birliktelik kuralları türetilir ve bu kurallarının her biri için güven ölçüleri belirlenir.

ERDEM , ÇİTPİT , DERYA için birliktelik kurallarının alt kümesi ;

{ ERDEM , ÇİTPİT } , { ERDEM , DERYA } , { ÇİTPİT , DERYA } , { ÇİTPİT } , { DERYA } , { ERDEM }

Birliktelik	Açıklama	Güven
ERDEM & ÇİTPİT -> DERYA	ERDEM ve ÇİTPİT in bulunduğu item-sette DERYA nın olma olasılığı	3/3=%100
ERDEM & DERYA -> ÇİTPİT	ERDEM ve DERYA nın bulunduğu item-sette ÇİTPİT in olma olasılığı	3/5=%60
ÇİTPİT & DERYA -> ERDEM	ÇİTPİT ve DERYA nın bulunduğu item-sette ERDEM in olma olasılığı	3/5=%60
ÇİTPİT -> ERDEM & DERYA	ÇİTPİT in bulunduğu item-sette ERDEM ve DERYA nın olma olasılığı	3/7 = %42
DERYA -> ERDEM & ÇİTPİT	DERYA nın bulunduğu item-sette ÇİTPİT ve ERDEM in olma olasılığı	3/8 = %38
ERDEM -> ÇİTPİT & DERYA	ERDEM in bulunduğu item-sette ÇİTPİT ve DERYA nın olma olasılığı	3/5 = %60

Apriori Algoritması(Örnek Çalışma)

ERDEM , DERYA , MAZLUM için birliktelik kurallarının alt kümesi ;

{ERDEM , DERYA} , {ERDEM , MAZLUM} , {DERYA , MAZLUM} , {DERYA} , {MAZLUM} , {ERDEM}

Birliktelik	Açıklama	Güven
ERDEM & DERYA -> MAZLUM	ERDEM ve DERYA bulunduğu item-sette MAZLUM un olma olasılığı	$3/5=60\%$
ERDEM & MAZLUM -> DERYA	ERDEM ve MAZLUM un bulunduğu item-sette DERYA nın olma olasılığı	$2/3=66\%$
DERYA & MAZLUM -> ERDEM	DERYA ve MAZLUM un bulunduğu item-sette ERDEM in olma olasılığı	$3/3=100\%$
DERYA -> ERDEM & MAZLUM	DERYA in bulunduğu item-sette ERDEM ve MAZLUM un olma olasılığı	$3/4=75\%$
MAZLUM -> ERDEM & DERYA	MAZLUM un bulunduğu item-sette DERYA nın ve ERDEM in olma olasılığı	$3/4=75\%$
ERDEM -> DERYA & MAZLUM	ERDEM in bulunduğu item-sette DERYA ve MAZLUM un olma olasılığı	$3/5=60\%$

Apriori Algoritması(Örnek Çalışma)

Birliktelik	Açıklama	Güven
ERDEM & ÇİTPIT -> DERYA	ERDEM ve ÇİTPIT in bulunduğu item-sette DERYA nın olma olasılığı	3/3=%100
ERDEM & DERYA -> ÇİTPIT	ERDEM ve DERYA nın bulunduğu item-sette ÇİTPIT in olma olasılığı	3/5=%60
ÇİTPIT & DERYA -> ERDEM	ÇİTPIT ve DERYA nın bulunduğu item-sette ERDEM in olma olasılığı	3/5=%60
ÇİTPIT -> ERDEM & DERYA	ÇİTPIT in bulunduğu item-sette ERDEM ve DERYA nın olma olasılığı	3/5=%60
Birliktelik	Açıklama	Güven
DERYA -> ERDEM & MAZLUM	DERYA bulunduğu item-sette ERDEM ve MAZLUM un olma olasılığı	3/5=%60
ERDEM & MAZLUM -> DERYA	ERDEM ve MAZLUM un bulunduğu item-sette DERYA nın olma olasılığı	2/3=%66
DERYA & MAZLUM -> ERDEM	DERYA ve MAZLUM un bulunduğu item-sette ERDEM in olma olasılığı	3/3=%100
DERYA -> ERDEM & MAZLUM	DERYA'nın bulunduğu item-sette ERDEM ve MAZLUM un olma olasılığı	3/4=%75
MAZLUM -> ERDEM & DERYA	MAZLUM un bulunduğu item-sette DERYA nın ve ERDEM in olma olasılığı	3/4= %75
ERDEM -> DERYA & MAZLUM	ERDEM in bulunduğu item-sette DERYA ve MAZLUM un olma olasılığı	3/5=%60

Bu iki birliktelik kuralında güven_{eşik}=80 değeri dikkate alınarak düzenleme yapılırsa;

DERYA & MAZLUM -> ERDEM = %100

ERDEM & ÇİTPIT -> DERYA = %100 'a göre aşağıdaki sonuçlar çıkarılır.

**DERYA ve MAZLUM un birlikte satıldığı alışverişte ERDEM in satılma olasılığı %100 'dür
ERDEM in ve ÇİTPIT in birlikte satıldığı alışverişte DERYA nın satılma olasılığı %100 'dür.**

Kaynaklar :

- Veri Madenciliği Yöntemleri, Yalçın Özkan 06'2008
- Veri Madenciliği ,Gökhan Silahtaroğlu 06'2008
- İstanbul Ticaret Üniversitesi Derğisi Veri Madencilği Modeller Ve Uygulama Alanları (Serhat ÖZEKES)

