

Veri Madenciliği Uygulamaları

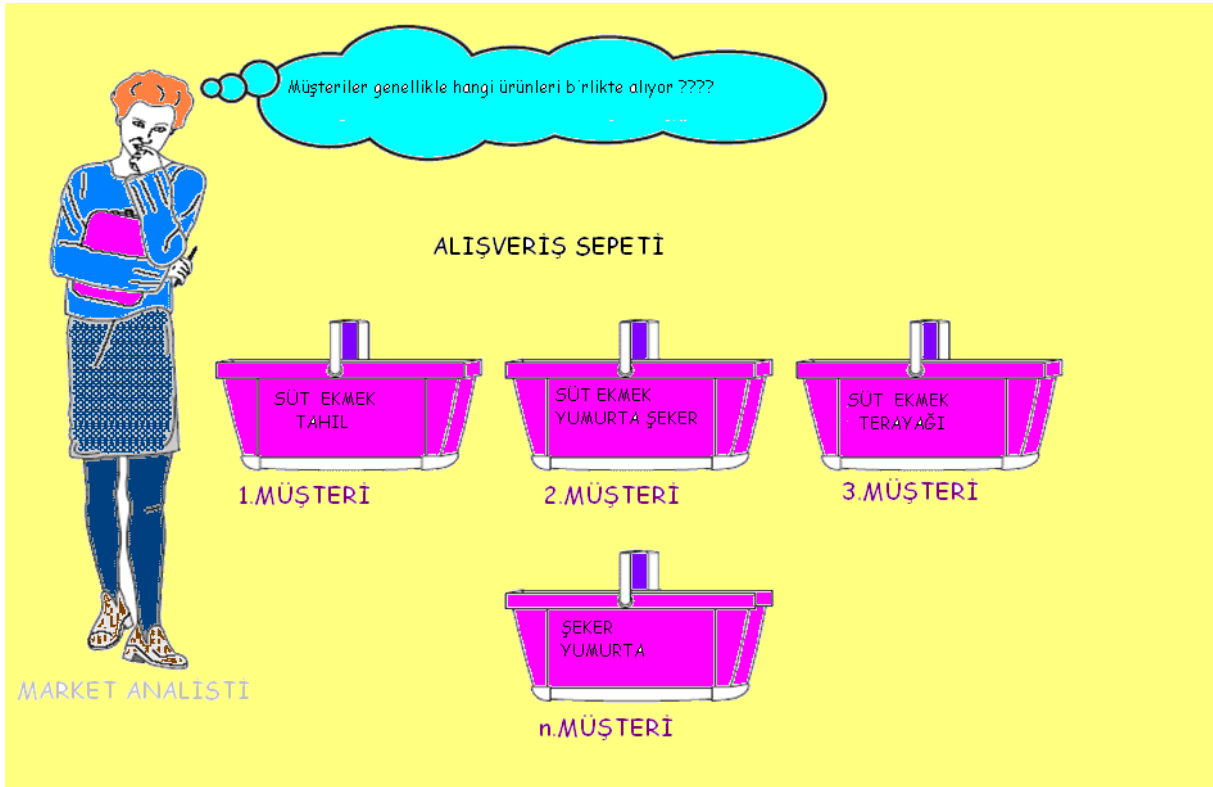
Hafta 9

Yrd. Doç.Dr. Nilüfer YURTAY

Birliktelik Kuralları ve İlişki Analizi

9.1 Giriş

Geçmiş tarihli hareketleri çözümlemek, karar destek sistemlerinde verilen kararın kalitesini artırmak için izlenen bir yaklaşımdır. 90'lı yılların başına değin teknik yetersizlikten dolayı, kurumlara veya müşterilere satış yapıldığı anda değil, belirli bir zaman aralığında (günlük, haftalık, aylık, yıllık) gerçekleşen satış hareketlerinin tamamına ilişkin genel veriler elektronik ortamda tutulmaktaydı. Barkot uygulamalarındaki gelişme ile, bir harekete ait verilerin satış hareketi olduğu anda toplanması ve elektronik ortama aktarılması olanaklı hale gelmiştir. Genellikle süpermarketlerin satış noktalarında bu tür veriler toplandığından, toplanan bu veriye **market sepeti verisi** adı verilmiştir. Market sepeti verisinde yer alan bir kayıta, tekil olan hareket numarası, hareket tarihi ve satın alınan ürünlere ilişkin ürün kodu, miktarı, fiyatı gibi bilgiler yer almaktadır. *Market sepet analizinde* (market basket analysis) amaç, satışlar arasındaki ilişkileri bulmak ve buna bağlı kuralları çıkarmaktır. Bu ilişkilerin bilinmesi, şirketin kârını arttırmak için kullanılabilir. Eğer X ürünü alanların Y ürünü de çok yüksek olasılıkla aldıkları biliniyorsa ve eğer bir müşteri X ürünü alıyor ama Y ürünü almıyorsa, o "**potansiyel bir Y müşterisidir**" denilebilir. Buna benzer veri analizleri yaparak her ürün için bir sonraki ayın satış tahminleri çıkarılabilir, birlikte satın alınan ürünler için promosyon uygulaması ve reyon dizilişleri yapılabilir, müşteriler satın aldıkları ürünlere göre gruplandırılabilir, yeni bir ürün için potansiyel müşteriler belirlenebilir.



9.2 Birliktelik Kuralları

Birliktelik kuralları (association rules), veri madenciliği alanında üzerinde çok fazla araştırma ve çalışma yapılmış olan ilgi çekici bir konudur. 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**. Bu tip birlikteliklerin keşfedilmesi,

müşterilerin hangi ürünleri bir arada aldıkları bilgisini ortaya çıkarır ve market yöneticileri de bu bilgi ışığında raf düzenlerini belirleyerek satış oranlarını artırabilir ve etkili satış stratejileri geliştirebilirler. Market sepeti çözümlemesinin son zamanlarda çok büyük ilgi ile karşılaşmasının sebebi kullanım kolaylığı ve anlaşılabilirliğidir. Market sepet analizi ile birliktelik kuralları çıkarımı ilk olarak Agrawal ve diğerleri tarafından 1993 yılında ele alınmıştır.

Kuralları oluşturabilmek için *destek* (support) ve *güven* (confidence) değerlerini kullanarak, kullanıcı tarafından belirlenmiş minimum destek ve minimum güven değerlerinden yaygın birlikteliklerin belirlenmesi amaçlanmıştır. Market sepet analizinde, nesneler müşteriler tarafından satın alınan ürünlerdir ve bir hareket (kayıt) birçok nesneyi içinde bulunduran tek bir satın almadır. Birliktelik kurallarının kullanışlı olması için hem konu ile ilgili hem de anlaşılabilir olması gerekir. Birliktelik kurallarında, kullanıcının kuralların tipini ve sayısını kontrol edebileceği çeşitli yollar vardır. En yaygın olarak kullanılan yöntem, eşik değerleri olarak bilinen minimum destek ve minimum güven değerlerinin belirlendiği yöntemdir. Bu yöntemde sadece kullanıcı tarafından belirlenen eşik değerlerinden büyük olan destek ve güven değerlerine sahip kurallar bulunur ve kullanılır. Diğer bir yöntemde kullanıcının sınırlanmış nesne tanımlamasıdır. Sınırlanmış nesne, kuralların içeriğinin sınırlanmasında kullanılan mantıksal bir ifadedir. Örneğin sınırlanmış nesne cips, kola ve hamburger olsun. Sadece cips, kola ve hamburger içeren kurallar ile ilgilenilir. Birliktelik kurallarındaki bir nesnenin ve bir işlemin tanımı uygulamaya bağlıdır. Market sepeti analizinde; nesneler, müşterilerin aldığı ürünler ve işlem, beraber alınan bütün nesnelerin kümesidir. Birliktelik kurallarında sıklıkla kullanılan birkaç önemli terim vardır. Bunlar; kuralın sol tarafını ifade eden önce (antecedent), kuralın sağ tarafını ifade eden sonuç (consequent), destek değeri, güven değeri, min_destek olarak gösterilen minimum destek değeri, min_güven olarak gösterilen minimum güven değeri, nesne küme, yaygın nesne kümesi ve aday nesne kümesidir.

X ürünü alan bir müşterinin Y ürününü de alma durumu(birliktelik kuralı) $X \rightarrow Y$ ile gösterilir.

Destek ölçütü

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

A ve B ürünlerinin birlikte satın alınma olasılığı güven değeridir. Güven değeri

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

Destek ve güven ölçütlerinin yanı sıra, bu değerleri karşılaştırmak için eşik değerlerine de ihtiyaç duyulmaktadır. Bulunan eşik değerlerinin, hesaplanan destek ve güven değerlerinden küçük olması beklenir. Hesaplanan destek ve güven değerlerinin büyüklük derecesi birliktelik kurallarının da o kadar güçlü olduğunu ifade eder.

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

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

Burada $X = \{Pantolon, Kazak\}$ 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:

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

Eğer pantolon ve kazak alanların sayısının 4 olduğunu farzedelim. Güven ölçütü

$$giiven(Pantolon, Kazak \rightarrow \text{Çorap}) = \frac{sayi(Pantolon, Kazak, \text{Çorap})}{sayi(Pantolon, Kazak)} = \frac{7}{4} = 1,75 \text{ olacaktır.}$$

Alışveriş yerleri genel olarak müşteri bilgileri ele geçirirler. Satılan her bir hareket sepet ("basket") olarak adlandırılır. Market –Sepet analizi, müşteri eğilimlerini tanımlayan sepet verilerini analiz eder.

Müşteri Numarası	Aldığı Çikolata Markası
MSNO101	ERDEM, ŞOKOKO, ÇİTPİT, DERYA
MSNO102	ENFES, DERYA, MAZLUM, ŞOKOKO
MSNO103	ERDEM, DERYA, ENFES
MSNO104	ŞOKOKO, ÇİTPİT, DERYA, POTPORİ,
MSNO105	ERDEM, DERYA, MAZLUM, ÇİTPİT
MSNO106	MİS, DERYA, ÇİTPİT
MSNO107	ERDEM, ZARİF, DERYA
MSNO108	MAZLUM, YURDUM, ÇİTPİT
MSNO109	YURDUM, ŞOKOKO, ÇİTPİT
MSNO110	ÇİTPİT, MİS, ERDEM, MAZLUM, DERYA

Birliktelik kuralları, item-setler arasındaki eğilimi ya da ilişkiyi bulur. Item set, itemlerin kümesini oluşturur. Her bir hareket, **item set** olarak adlandırılır. Örneğin MSNO108 numaralı müşterinin yapmış olduğu alışverişteki “MAZLUM-YURDUM-ÇİTPİT” bir item settir.

9.3 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 sayıları hesaplanır.Eşik değeri ile karşılaştırılan destek değerlerinin içinden eşik değerinden düşük olanlar çıkarılır.*
- *Kalan ürünler ikişerli gruplanarak, grup destek sayıları 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.*

Örnek Çalışma

Müşteri Numarası	Aldığı Çikolata Markası
MSNO101	ERDEM, ŞOKOKO, ÇİTPİT, DERYA
MSNO102	ENFES, DERYA, MAZLUM, ŞOKOKO
MSNO103	ERDEM, DERYA, ENFES
MSNO104	ŞOKOKO, ÇİTPİT, DERYA, POTPORİ,
MSNO105	ERDEM, DERYA, MAZLUM, ÇİTPİT
MSNO106	MİS, DERYA, ÇİTPİT
MSNO107	ERDEM, ZARİF, DERYA
MSNO108	MAZLUM, YURDUM, ÇİTPİT
MSNO109	YURDUM, ŞOKOKO, ÇİTPİT
MSNO110	ÇİTPİT, MİS, ERDEM, MAZLUM, DERYA

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

Destek_{eşik} = %30

Güven_{eşik} = %80

Burada eşik destek sayısı $0,30 * 10 = 3$ dur.

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

Çikolata Markası	Destek Değeri
ERDEM	5
ŞOKOKO	4
ÇİTPİT	7
MAZLUM	4
ENFES	2
DERYA	8
POTPORI	1
MİS	2
YURDUM	2
ZARIF	1

Çikolata Markası	Destek Değeri
ERDEM	5
ŞOKOKO	4
ÇİTPİT	7
MAZLUM	4
DERYA	8

- **Kalan ürünler ikiyeşerli gruplanarak, grup destek sayıları hesaplanır. Tekrar eşik değerleri ile karşılaştırılan destek değerlerinden eşik değerinin altında kalanlar iptal edilir.**

Çikolata Markası	Destek Değeri
ERDEM , ŞOKOKO	1
ERDEM , ÇİTPİT	3
ERDEM , MAZLUM	2
ERDEM , DERYA	5
ŞOKOKO , ERDEM	1
ŞOKOKO , ÇİTPİT	3
ŞOKOKO , MAZLUM	2
ŞOKOKO , DERYA	3
ÇİTPİT , ERDEM	3
ÇİTPİT , ŞOKOKO	3
ÇİTPİT , MAZLUM	3
ÇİTPİT , DERYA	5
MAZLUM , ERDEM	2
MAZLUM , ŞOKOKO	2
MAZLUM , ÇİTPİT	3
MAZLUM , DERYA	3
DERYA , ERDEM	5
DERYA , ÇİTPİT	5
DERYA , MAZLUM	3
DERYA , ŞOKOKO	3

1 destek değerli 1 item set var,
2 destek değerli 2 item set var,
3 destek değerli 5 item set var,
5 destek değerli 2 item set var,

Çikolata Markası	Destek Değeri
ERDEM , ÇİTPİT	3
ERDEM , DERYA	5
ŞOKOKO , ÇİTPİT	3
ŞOKOKO , DERYA	3
ÇİTPİT , MAZLUM	3
ÇİTPİT , DERYA	5
MAZLUM, DERYA	3

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

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

Üçlü birlikteliklerin destek değerleri dikkate alınırsa , destek 3 olarak seçilebilir.

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

- 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

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

Bu iki birliktelik kuralında $\text{güven}_{\text{eşik}}=80$ değeri dikkate alınarak düzenleme yapılırsa;

DERYA & MAZLUM \rightarrow ERDEM = %100

ERDEM & ÇİTPİT \rightarrow DERYA = %100

'a göre aşağıdaki sonuçlar çıkarılır.

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