



Research Article

ASSOCIATION RULE MINING AND INTERESTINGNESS MEASURES: A CASE STUDY

BİRLİKTELİK KURALLARI MADENCİLİĞİ VE İLGİNÇLİK ÖLÇÜMLERİ: BİR VAKA ÇALIŞMASI

Başar KARASU¹ | Onur DOĞAN^{2,*}

¹ Yüksek Lisans Mezunu, Dokuz Eylül Üniversitesi, Türkiye, ORCID: 0000-0001-2345-6789

² Doç. Dr., İzmir Meslek Yüksekokulu, Dokuz Eylül Üniversitesi, Türkiye, onur.dogan@deu.edu.tr, ORCID: 0000-0002-7916-8994

Article Info:

Received : October 16, 2020

Revised : December 10, 2020

Accepted : December 24, 2020

Keywords:

data mining

association rules

interestingness measures

Anahtar Kelimeler:

veri madenciliği

birlikte lık kuralları

ilginçlik ölçüleri

Note:

Bu makale Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Yönetim Bilimi Anabilim Dalında sunulmuş bir yüksek lisans teziden türetilmiştir.

DOI: 10.46238/jobda.811464

ABSTRACT

Association Rule Mining is a method which used in data mining. In this method, the objects that are seen together in the data sets are identified and interesting patterns that will help the decision maker are revealed. Revealed associations are expressed in the form of rules. Interestingness measures are used to evaluate the rules. The decision-maker pursues strong and interesting patterns and uses this analysis in the most efficient and correct manner on behalf of the firm.

In this study, the interestingness measures which are used in the evaluation of association rules are discussed extensively. The main purpose of the study is to contribute to the Turkish literature by presenting a large number of interesting measures used in the evaluation of association rules. In addition, some association rules have been obtained according to some criteria on the data of various stores of a firm and these rules have been evaluated by the interestingness measures. It can be stated that the findings are useful for the firm's decision makers.

ÖZET

Birlikte lık Kural Madenciliği, veri madenciliğinde kullanılan yöntemlerden biridir. Bu yöntemde, veri kümelerinde birlikte görülen nesnelere belirlenerek, karar vericiye yardımcı olacak ilginç örüntüleri ortaya çıkartılmaktadır. Ortaya konulan birliktelikler, kurallar biçiminde ifade edilir. Elde edilen kuralların değerlendirilmesi için ilginçlik ölçülerinden yararlanır. Karar verici, güçlü ve ilginç örüntülerin peşinden gider ve firması adına en verimli ve doğru şekilde bu analizden faydalanır.

Bu çalışmada, birliktelik kurallarının değerlendirilmesinde kullanılan ilginçlik ölçümleri kapsamlı bir biçimde ele alınmıştır. Çalışmanın ana amacı birliktelik kurallarının değerlendirilmesinde kullanılan çok sayıda ilginçlik ölçümünü sunarak, Türkçe yazına katkı sağlamaktır. Bununla birlikte, bir firmaya ait çeşitli şubelerin verileri üzerinden bir takım kriterler doğrultusunda bazı birliktelik kuralları elde edilmiş ve bunlar ilginçlik ölçümleri ile değerlendirilmiştir. Elde edilen bulguların firmanın karar vericileri için yararlı nitelikte olduğu belirtilebilir

© 2020 JOBDA All rights reserved

1 | GİRİŞ

Günümüz insanının yapmış olduğu her alışverişte, her telefon aramasında ve her bankacılık işleminde, uydulardan ve uzaktan algılayıcılardan elde edilerek kaydedilen, devlet ve işletme yönetimi altında

gerçekleştirilen işlemler neticesinde depolanan veriler her geçen an olağanüstü boyutlarda artış göstermektedir (Akpınar, 2000).

İşletmeler, müşterileri ile ilgili büyük miktarda, farklı özellik ve niteliklere sahip veriye sahiptir.

*Corresponding author.

E-Posta: onur.dogan@deu.edu.tr (O. Doğan)

İşletmelerin sahip oldukları bu veriyi en etkin biçimde kullanabilmeleri verinin işletme tarafından doğru anlaşılabilmesi ile mümkündür. Bu çabanın ve beraberinde elde edilen bilginin ise işletmelerin rekabet güçleri açısından önemi açıktır. Fazla sayıda şube ile çok müşteriye hizmet veren sektörlerde oluşturulan veri tabanlarından elde edilecek bilgi işletmenin rekabet gücünü arttıracaktır. Veri madenciliği tekniklerine duyulan ihtiyaç bu noktada kendini göstermektedir. Bu teknikler büyük hacimli ve insan eliyle bilginin elde edilmesinin zor olduğu veri kümelerinden; yararlı ve kullanışlı bilgilerin daha hızlı ve kolay elde edilmesinden kullanılır. Sınıflama, kümeleme, birliktelikleri kuralları, tahminleme, örüntü bulma, vb. birçok metot, veri madenciliği başlığı altında, işletmeler için kullanışlı olabilecek bilgileri ortaya koymayı ve işletmelerin sahip olduğu veri kümeleri içerisindeki gizli bilgileri açığa çıkarmayı amaçlamaktadır.

Birliktelik kuralları madenciliği de veri madenciliği kapsamında en çok kullanılan metotlardan biridir. Bu kapsamda kullanılan algoritmalar; veri kümelerindeki hangi nesnelerin birlikte olduğunun ortaya çıkarılmasında, ilginçlik ölçüleri ise; bu birlikteliklerin değerlerinin ortaya konulmasında kullanılmaktadır. Birliktelik kurallarında kullanılan algoritmalar ve ilginçlik ölçümleri değişkenlik göstermektedir. Araştırmacılar farklı metotlar izleyerek çok sayıda algoritma ve ilginçlik kural ölçüsü geliştirmiştir.

Bu çalışmanın temel amacı birliktelik kurallarının değerlendirilmesinde kullanılan ilginçlik ölçümlerinin kapsamlı bir şekilde sunulmasıdır. Ayrıca, uygulama kısmında perakende sektöründe faaliyet gösteren firmanın bir şubesine ait satış verilerinden yola çıkarak elde edilen bulgulara yer verilecektir.

2 | VERİ MADENCİLİĞİ KAVRAMI

Veri Madenciliği, büyük miktarda veri yığınlarının anlamlı ve faydalı bilgilere dönüştürülmesi sürecidir. Sahip olunan veri kümelerinin çok büyük ve karmaşık olması, klasik matematiksel ve istatistiksel metotları kullanarak verilerin çözülmesi konusunda yetersiz kalabilmektedir. Zira büyük ve karmaşık veri kümeleri üzerinde yapılan analizler yoğun vakit ve iş yüküne sebebiyet vermekte çoğu zaman da uygun çözümlere ulaşılmaması konusunda başarısızlıklar söz konusu olmaktadır.

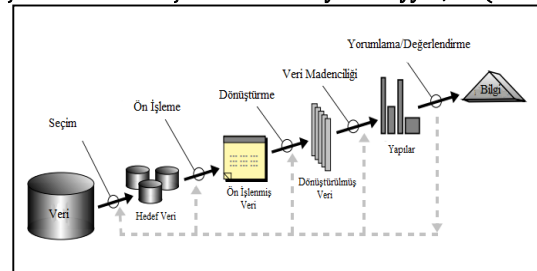
Veri madenciliği, büyük veri kümelerinin, istatistik, matematik ya da örüntü tanımlama teknikleri yardımıyla incelenerek bu veriden yeni ilişkilerin, örüntülerin ve trendlerin keşfedilmesi sürecidir (Larose, 2005).

Veri Madenciliği klasik istatistiksel çözümlerden farklı olarak, daha fazla değişken ve daha fazla gözleme sahip veri kümelerini analiz etmek için kullanılmaktadır. Veri Madenciliği sahip olduğu algoritmaları kullanarak karmaşık veri kümelerini analiz edip, faydalı bilgiler ve modeller keşfetmeye yarayan bir süreçten oluşur. Bu süreçte Veri Tabanı Sistemleri, Yüksek Performanslı Programlama, Veri Görselleştirme, Görüntü ve Sinyal İşleme, İstatistik, Yapay Öğrenme ve Örüntü Tanıma gibi disiplinler arası kavramlar kullanılır (Ledoelter, 2013, Zaki ve Meira, 2014).

Veri madenciliği büyük miktarda veri içinden, gelecek ile ilgili tahmin yapılmasını sağlayacak bağıntı ve kuralların bilgisayar programları kullanılarak aranmasıdır. Bu durum; veri analizi yapılarak, bir mal için bir sonraki ayın satış tahminleri ile yapıp, müşteriler satın aldıkları mallara bağlı olarak gruplandırılıp, yeni bir ürün için potansiyel müşteriler belirlenip, müşterilerin zaman içindeki hareketleri incelenerek onların davranışları ile ilgili tahminler yapılabilmesidir. Binlerce mal ve müşterin var olduğu dikkate alındığında bu analizlerin gözle ve elle yapılamayacağı, otomatik olarak yapılmasının gerektiği ortaya çıkar ve veri madenciliği bu noktada devreye girer (Alpaydın, 2000).

Veri madenciliği ve veri tabanlarında bilgi keşfi (VTBK) süreci kavramları birbirinin yerine kullanılabilir ve VTBK sürecinin bir adımıdır (Moen, 2005). VTBK süreç modeli Şekil 1'de gösterildiği gibidir:

Şekil 1: VTBK Süreci Basamakları Kaynak: Fayyad, vd. (1996)



Görüldüğü üzere VTBK süreci verinin saklandığı ortamlarından uygun verilerin seçilmesi ile başlayan, ön işleme, dönüştürme, analizlerin yapılması ve değerlendirme basamaklarından sonra veriden bilgi elde edilmesine uzanan bir süreçtir.

Raporlama araçlarının günümüz veri yığınları karşısında yetersiz kalması, VTBK'nin ortaya çıkmasına neden olmuştur. VTBK süreci içerisinde Veri Madenciliği adımı, sürecin en önemli kısmını oluşturmaktadır, zira bu aşamada toplanmış verinin analizi için en uygun modelin kurulması ve değerlendirilmesi aşamaları yer almaktadır.

3 | BİRLİKTELİK KURALLARI MADENCİLİĞİ

Veri madenciliği modelleri tahminleyici ve tanımlayıcı modeller olmak üzere iki başlıkta sınıflandırılmaktadır. Tahminleyici modeller; sınıflandırma, regresyon, zaman serisi analizleri, öngörüleme ve tanımlayıcı modeller; kümeleme, özetleme, birliktelik kuralları, dizi keşfi olarak sıralanabilir (Dunham, 2006). Birliktelik analizi, veri kümesinde birlikte görülen nesnelere kümelerini ve bu kümeler arasındaki ilişkileri etkili bir şekilde keşfetmek için kullanılır. Keşfedilen örüntüler, sık gözlenen nesne kümeler ya da kurallar olarak gösterilir. Birliktelik analizinde bu örüntülerin ilginç olması beklenir ve örüntülerin ilginçliğinin belirlenmesi için çeşitli ölçümler sunulmuştur. Bu ölçümlerin seçimi ise hassas bir konudur. Veri kümesinin büyüklüğü arttıkça ilginç örüntülerin belirlenmesi daha da zorlaşır (Tan, vd. 2006) Birliktelik kurallarını çıkarma işi büyük bir işlemsel veritabanı içinde bir kerde tarama yapmayı gerektirdiğinden, işlem gücünün muazzam seviyede olması zorunludur (Lee ve Siau, 2001). Veritabanı içinde yer alan kayıtların birbirleriyle olan ilişkilerini inceleyerek, hangi olayların eş zamanlı olarak birlikte gerçekleşebileceklerini ortaya koymaya çalışan veri madenciliği yöntemleri bulunmaktadır. Bu ilişkilerin belirlenmesiyle "birliktelik kuralları" elde edilir (Silahtaroglu, 2008).

3.1 | Temel Kavramlar ve Kural Çıkarımı

Birliktelik kural analizleri değişkenlerden hangilerinin "birlikte" olduğunun bulunmasına yöneliktir. İşletme dünyasında daha yaygın olarak "benzeşim analizleri" ya da "Pazar sepeti analizi" adları ile de bilinen birliktelik analizleri iki ya da daha fazla değişken arasındaki tespit edilmemiş kuralların araştırılması üzerinedir.

Birliktelik kuralları; "EĞER -öncül ifade-, İSE -ardıl ifade-"formunda, destek ve güven ölçümleri olan kurallardır (Larose, 2005). Birliktelik Analizi olarak bilinen yöntem, büyük veri kümelerindeki ilginç ilişkilerin keşfedilmesinde kullanılmaktadır. Ortaya çıkan ilişkiler "birliktelik kuralları" olarak isimlendirilmektedir (Tan, vd. 2006). Kuralın, verilen karar öncesine kadar olan kısmına öncül denir. Kuralın karar kısmına ise ardıl adı verilir. Destek (support) ve güven (confidence) değeri, bir birliktelik kuralının gücünü gösteren ölçü birimleri olarak değerlendirilebilir.

Birliktelik kuralları için birer ilginçlik ölçüsü olan güven ve destek değerleri aşağıdaki gibi gösterilir (Agrawal, vd. 1993);

D: Veri,

ti: Verideki kayıtlar, $D=\{t_1, t_2, \dots, t_i, \dots, t_n\}$

X, Y: Kurallardaki öğeler (Öncül ve Ardıl)

$X \rightarrow Y$, X'in öncül, Y'nin ardıl olduğu bir kural ($X \cap Y = 0$) olmak üzere;

Destek değeri (dd):

$$dd(X) = \frac{|t \in D, X \subset t|}{|D|}$$

Güven değeri (gd):

$$gd(X \rightarrow Y) = \frac{dd(X \rightarrow Y)}{dd(X)}$$

Destek değeri veri içindeki tekrar eden kayıtların, tüm veriye oranı iken güven değeri ise bir kuralın destek değerinin, öncül öğelerin destek değerine oranıdır.

3.2 | İlginçlik Ölçü Birimleri

Birliktelik analizleri sonucunda oluşan yüksek destek ve güven değerlerine sahip kurallar, güçlü kurallar olmasına rağmen her zaman ilişkinin gerçek gücünü göstermeyebilir. Örneğin bir kanser hastasına verilen A ilacının B ilacıyla birleştirilerek kullanılması gerekiyor ise, A ile B arasında yüksek bir destek ve güven değeri çıkacaktır. Ancak A ile B arasında oluşturulabilecek kuralın ilginç bir kural olmadığı aşikârdır A ilacı ile B ilacının birleştirilerek kullanıldığını bilmeyen bir analizci için bu tam aradığı bir kural olarak değerlendirilebilir. Ancak analizi yapan bu işin uzmanı bir doktor ise bu kuralın ilginç olmadığını ve değerlendirme dışında tutulması gerektiğini söyleyebilecektir. Buna benzer kurallar araştırmacıyı yanıltacak yapıdadır. Birliktelik kurallarının daha anlamlı kurallardan oluşturulması için ilginçlik ölçü birimlerinin de kullanılması gerekmektedir. İlginçlik ölçü birimlerinin kullanılmadan elde edilen kurallar, o işin uzmanı olunmadığı sürece de anlamlı olarak görülebilecektir. Özel ölçümleri belirlemek için gerekli uzman bilgisine ulaşmak genellikle kolay değildir ya da uzman bilgisi güvenilir olmayabilir. Ayrıca, bu uzman bilgisine ulaşmanın maliyeti de yüksektir (McGarry, 2005). Buradan hareketle destek ve güven değerinin yanı sıra birliktelik kurallarının genelliği ve güvenilirliğini değerlendirmek için olasılık temelli objektif ölçümler pek çok araştırmacı tarafından kapsamlı olarak incelenmiştir. Farklı çalışmalardan derlenen ilginçlik ölçü birimleri izleyen kısımdaki gibidir.

Laplace Correction

$$\text{Laplace Correction } (M \rightarrow K) = \frac{dd(X \rightarrow Y)+1}{dd(X)+2}$$

eşitliği ile ifade edilir.

Güven değeri, [0, 1] aralığında değer alırken Laplace Correction değeri [0, 1) aralığında değer almaktadır (Azevedo ve Jorge, 2007).

Bütünlük Değeri

Güven değeri, öncül ve ardılın birlikte olduğu kayıt sayısının öncül kayıt sayısına oranıyken ($X \cap Y = 0$), bütünlük değeri, ardılın kayıt sayısına oranıdır (Bramer, 2007):

$$\text{Bütünlük Değeri } (X \rightarrow Y) = \frac{dd(X \cup Y)}{dd(Y)}$$

şeklinde gösterilir.

Lift

Lift değeri şu şekilde hesaplamak mümkündür (Zhao, vd. 2009):

$$\text{Lift}(X \rightarrow Y) = \frac{gd(X \cup Y)}{dd(Y)}$$

Lift değerini hesaplayan bir başka formül ise şu şekildedir (Brin, vd. 1997):

$$\text{Lift}(X \rightarrow Y) = \frac{dd(X \cup Y)}{dd(X).dd(Y)}$$

Lift değerinin 1'den küçük olması X ile Y arasında negatif korelasyon olduğuna, 1'den büyük olması pozitif korelasyona olduğuna, 1'e eşit olması ise X ve Y'nin birbirinden bağımsız olduğunu göstermektedir. Lift değerinin 1'e eşit olduğu durumda, bu kuralların ilginç kurallar olmadığı, ilginç kuralların her iki değişkenin birbirine negatif ya da pozitif yönde bağımlı olmaları gerektiği yargısına varılabilmektedir. Bir başka deyişle bağımlı ve birbiri ile ilişkili kurallar ilginçtir (Zhao, vd. 2009; Brin, vd. 1997).

Leverage

Piatetsky-Shapiro olarak da adlandırılan Leverage aşağıdaki gibi tanımlanmıştır (Kannan ve Bhaskaran, 2009).

$$\text{Leverage } (X \rightarrow Y) = gd(X \cup Y) - [dd(Y).dd(X)]$$

Bu simetrik ölçüm, bir işlemde meydana gelen X U Y'nin gerçek olasılığı yani öncül ve ardılın güven değeri ile X ve Y'nin istatistiksel olarak bağımsız olma olasılığı arasındaki farktır (Zaki ve Meira, 2014). Minimum değer -1, maksimum değer 1 olması ardıl ve öncüllerin birbiriyle bağımlı olduğunu, 0 olması ise ardıl ve öncüllerin birbirinden bağımsız olduğunu göstermektedir.

Conviction

Conviction değeri aşağıdaki gibi hesaplanır (Brin, vd. 1997):

$$\text{Conviction}(X \rightarrow Y) = \frac{1 - dd(Y)}{1 - gd(X \cup Y)}$$

Conviction değerinin 1'den küçük olması negatif korelasyona, Conviction değerinin 1'den pozitif korelasyona, Conviction değerinin 1'e eşit olması ise bağımsızlığa işaret etmektedir. Lift değerinin aksine conviction değeri tek yönlü değildir (değişme özelliği

yoktur), yani simetrik bir ölçüm değildir (Jimenez, 2013).

Kesinlik Faktörü

Kesinlik Faktörü, birliktelik kurallarındaki belirsizliği ölçmek için geliştirilmiştir ve aşağıdaki formül ile gösterilir (Berzal, vd. 2002):

$$\text{Kesinlik Faktörü } (X \rightarrow Y) = \frac{gd(X \cup Y) - dd(Y)}{dd(Y)}$$

Certainty Factor, $M \rightarrow K$ gibi bir kuralda, yalnızca öncülün (M) olduğu işlemleri dikkate alarak, kuralın ardıl (K) içerdiği olasılığı ölçer. Bu ölçü -1 ile 1 arasında bir değer aldığı anda öncül ve ardılar birbiriyle bağımlı, 0 olduğunda ise öncül ve ardılın birbirinden bağımsız olduğunu göstermektedir.

Klosgen

Klosgen değeri, aşağıdaki formüldeki gibi tanımlanır (Tan, vd. 2004)

$$\text{Klosgen } (X \rightarrow Y) = \left[\sqrt{dd(X \cup Y)} \cdot gd(X \rightarrow Y) \right] - dd(Y)$$

Klosgen değeri, -1 ile 1 arasında bir değer aldığı anda öncül ve ardıl arasında bağımlı bir ilişki, 0 değerini aldığı anda öncül ve ardıl arasında bağımsız bir ilişki vardır.

Cosine

Cosine değerinin aşağıdaki gibi hesaplanır (Tan ve Kumar, 2000)

$$\text{Cosine } (X \rightarrow Y) = \frac{dd(XUY)}{\sqrt{dd(X).dd(Y)}}$$

Bu değer 0 ile 1 arasında bir değer alır. Payda değeri 0 olamayacağı için Cosine değeri öncül ve ardıl arasında bağımlı bir ilişki olduğunu ortaya koyacaktır. Bu ilişkinin ilginçlik gücü 0'a yaklaştıkça azalacak, 1'e yaklaştıkça artacaktır (Tan ve Kumar, 2000).

Jaccard Değeri

Jaccard değeri, aşağıdaki gibi gösterilmektedir (Rijsbergen, 2979):

$$\text{Jaccard } (X \rightarrow Y) = \frac{dd(XUY)}{dd(X) + dd(Y) - dd(XUY)}$$

Jaccard değeri, öncül ve ardılar arasındaki ilişkinin 0 ile 1 arasında bir değer aldığını gösterir. Değerin yüksek olması ilişkinin ilginçlik gücünün de yüksek olduğunu gösterir.

Gini İndeks

Gini indeks değeri, aşağıdaki formülasyon ile gösterilmektedir (Shaikh, 2013):

$$\text{Gini İndeks}(X \rightarrow Y) = dd(X)[gd(X \rightarrow Y)^2 + gd(X \rightarrow Y')^2] + dd(Y').[gd(X' \rightarrow Y)^2 + gd(X' \rightarrow Y')^2] - dd(Y)^2 - dd(Y')^2$$

J-Ölçüsü

J-Ölçüsü değeri, aşağıdaki gibi ifade edilmektedir (Smyth ve Goodman, 1991):

$$J\text{-Ölçüsü } \frac{(X \rightarrow Y)}{gd(X \rightarrow Y)} = \frac{dd(X \rightarrow Y)}{gd(X \rightarrow Y)} \log \left[\frac{dd(X \rightarrow Y)}{dd(Y)} \right] + \frac{dd(X \rightarrow Y)}{gd(X \rightarrow Y)} \log \left[\frac{dd(X \rightarrow Y)}{dd(Y')} \right]$$

Bu ölçü değeri simetrik değildir. 0 değeri öncül ve ardılların birbiriyle bağımsız olduğunu gösterir.

Yule's Q

Bazı ilginçlik ölçümleri, ilişkiler arasındaki eksikliklerin dikkate alınmasını desteklemektedir. Yule's Q değeri "Odds Ratio (Olasılık Oranı)" değerini temel alan böyle bir ölçümdür. Ancak Odds Ratio'nun aksine, Yule's Q; $P(A) > 0$ ve $P(B) > 0$ olan durumlarda tanımlanmıştır (Shaikh, 2013).

$$Yule's Q(X \rightarrow Y) = \frac{[dd(X \rightarrow Y)dd(X' \rightarrow Y')] - [dd(X' \rightarrow Y)dd(X \rightarrow Y')]}{[dd(X \rightarrow Y)dd(X' \rightarrow Y')] + [dd(X' \rightarrow Y)dd(X \rightarrow Y')]}$$

Yule's Q değeri öncül ve ardılın birbirinden bağımsız olması durumunda 0 değerini alır.

Yule's Q değerinin pozitif olması; öncül ve ardılın birlikte olma olasılığının daha yüksek olduğunu, negatif olması ise daha düşük olduğunu göstermektedir (Shaikh, 2013).

Yule's Y

Birleşme katsayısı (The Coefficient of Colligation) olarak da bilinen Yule's Y değeri, aşağıdaki formülasyon ile gösterilmektedir (Shaikh, 2013).

$$Yule's Y(M \rightarrow K) = \frac{\sqrt{dd(X \rightarrow Y).dd(X' \rightarrow Y')}}{\sqrt{dd(X).dd(Y)}}$$

Yule's Y değeri, +1 ile -1 arasında değişen bir değerdedir. Pozitif değer, pozitif korelasyonu, negatif değer ise negatif korelasyonu gösterir. Değerin 0 olması ise öncül ve ardıl arasında hiç ilişki olmadığını yani birbirlerinden bağımsız olduğunu gösterir.

Odds Oranı

Odds Ratio (olasılık oranı) ilginçlik ölçümleri içinde kullanılabilecek bir başka ölçüm değeridir ve aşağıdaki formül ile gösterilir (Tan, vd. 2006).

$$Odds Ratio (M \rightarrow K) = \frac{dd(X \rightarrow Y).dd(X' \rightarrow Y')}{dd(X \rightarrow Y').dd(Y \rightarrow X')}$$

Odds Ratio değerinin 1 e yaklaşması öncül ve ardıl arasındaki ilişkinin ilginçlik gücünün de yüksek olduğunu göstermektedir. Değer 0 ise, aralarında bir ilişki yoktur.

Müşterek Güç

Müşterek Güç (Collective Strength) değeri aşağıdaki gibi hesaplanır (Tan, vd. 2006).

$$Müşterek Güç (X \rightarrow Y) = \frac{dd(X \rightarrow Y) + gd(X' \rightarrow Y')}{\sqrt{[dd(X).dd(Y)] + [d(X').d(Y')]}}$$

Collective Strength değeri öncül ve ardıl arasındaki ilişkinin ilginçlik gücünü gösterir. Değer 1'e yaklaştıkça, ilişkinin ilginçlik gücü de artar.

Doğrusal Korelasyon Katsayısı

Doğrusal Korelasyon Katsayısı (DKK) aşağıdaki gibi hesaplanır (Kirchgesner, vd. 2016).

$$DKK(X \rightarrow Y) = \frac{dd(X \rightarrow Y) + [dd(X).dd(Y)]}{\sqrt{dd(X).dd(Y).dd(X').dd(Y')}}}$$

Loevinger Değeri

Loevinger değeri aşağıdaki gibi hesaplanır (Geng ve Hamilton 2006).

$$Loevinger (X \rightarrow Y) = \frac{1 - [dd(X).dd(Y')]}{dd(X \rightarrow Y')}$$

Loevinger değeri "1 - [dd(M).dd(K')]" pozitif bir değer olacağından, 0 ile 1 arasında yani pozitif bir değer olacaktır ve bu ilişkinin ilginçlik gücü 0 dan uzaklaştıkça artacaktır.

Information Gain

Bu değer aşağıdaki gibi hesaplanır (Masood ve Ouaguenouni, 2013).

$$Information Gain (X \rightarrow Y) = \log \frac{dd(X \rightarrow Y)}{dd(X).dd(Y)}$$

Sebag - Schoenauer Değeri

Sebag-Schoenauer değeri aşağıdaki gibi hesaplanır (Sebag ve Schoenauer, 1998).

$$Sebag-Schoenauer (X \rightarrow Y) = \frac{dd(X \rightarrow Y)}{dd(X \rightarrow Y')}$$

Sebag-Schoenauer değeri öncül ve ardıl arasındaki ilişkinin ilginçlik gücünü gösterir. Değer 1'e yaklaştıkça, ilişkinin ilginçlik gücü de artar. Değerin 0 olması durumunda öncül ve ardılın birbirinden bağımsız olduğunu göstermektedir.

Least Contradiction

Bu değer aşağıdaki gibi hesaplanır (Aze ve Kodratoff, 2002).

$$Least Contradiction (X \rightarrow Y) = \frac{dd(X \rightarrow Y) - dd(X \rightarrow Y')}{dd(Y)}$$

Bu değer, 1 ile -1 arasında yer alır. Değerin yüksek olması ilişkinin ilginçlik gücünün de yüksek olduğunu, değer 0 olması ise aralarında bağımsız bir ilişki olduğunu gösterir.

Örnek ve Karşı-Örnek Oranı

Bu değer aşağıdaki gibi hesaplanır (Gras, vd. 1996).

$$\frac{\text{Örnek ve Karşı-Örnek Oranı}}{(X \rightarrow Y)} = \frac{1 - dd(X \rightarrow Y')}{dd(X \rightarrow Y)}$$

Örnek ve Karşı Örnek Oranı (Example and Counter-Example) pozitif bir değer olacaktır. Değer 0'dan uzaklaştıkça öncül ve ardıl arasındaki ilişkinin ilginçlik gücü de artar.

Leverage-2

Leverage-2 değeri aşağıdaki gibi hesaplanır (Piatetsky-Shapiro, 1991):

$$\text{Leverage-2 } (X \rightarrow Y) = dd(X \rightarrow Y) - [dd(X) \cdot dd(Y)]$$

Leverage-2 değerinin 0 olması durumunda öncül ve ardıllar arasında bağımsız bir ilişki vardır. Değer 0'dan uzaklaştıkça ilişkinin ilginçlik gücü de artar.

Coherence

Coherence değeri aşağıdaki gibi hesaplanır (Kannan ve Bhaskaran, 2009).

$$\text{Coherence } (X \rightarrow Y) = \frac{dd(X \rightarrow Y)}{[dd(X) + dd(Y) - dd(X \rightarrow Y)]}$$

Coherence değeri 0'dan büyük yani pozitif bir değer alır. Değer 0'dan uzaklaştıkça ilişkinin ilginçlik gücü de artar.

Specificity-2

Specificity-2 değeri aşağıdaki gibi hesaplanır (Kannan ve Bhaskaran, 2009).

$$\text{Specificity-2 } (X \rightarrow Y) = gd(Y' \rightarrow X')$$

Specificity-2 değeri 0'dan büyük yani pozitif bir değer alır. Değer 0'dan uzaklaştıkça ilişkinin gücü de artar.

Kulczynski Değeri

Kulczynski değeri aşağıdaki gibi hesaplanır (Kannan ve Bhaskaran, 2009).

$$\text{Kulczynski } (X \rightarrow Y) = \frac{gd(Y \rightarrow X) + gd(X \rightarrow Y)}{2}$$

Kulczynski değeri, incelenen kuralda her iki işleme ait güven değerleri toplamının ortalamasıdır. Bu değer, pozitif bir değer olacaktır. Değer 0'dan ne kadar uzak ise ilişkinin ilginçlik gücü de o kadar yüksektir.

Tek Yönlü Destek Değeri

Bu değer aşağıdaki gibi hesaplanır (Kirchessner, vd. 2016).

$$\text{Tek Yönlü Destek } (X \rightarrow Y) = gd(X \rightarrow Y) \times \log_2 \frac{dd(X \rightarrow Y)}{dd(X) \cdot dd(Y)}$$

İki Yönlü Destek Değeri

İki yönlü destek değeri aşağıdaki gibi hesaplanır (Kirchessner, vd. 2016).

$$\text{İki Yönlü Destek } (X \rightarrow Y) = dd(X \rightarrow Y) \times \log_2 \frac{dd(X \rightarrow Y)}{dd(X) \cdot dd(Y)}$$

4 | VAKA ANALİZİ

Bu kısımda perakende sektöründe faaliyet gösteren çok şubeli bir firmaya ait bir şubenin, belirlenmiş kriterler doğrultusunda satış verilerinden yola çıkarak gerçekleştirilen bir uygulamaya yer verilecektir. Uygulama aşamasında, elde edilen veriler belirli işlemlerden geçirildikten sonra, WEKA programı kullanılarak birliktelik kural madenciliği analizlerine tabi tutularak anlamlı kurallar elde edilmeye çalışılmıştır.

4 1 | Veri Seçimi

Uygulamada, hızlı tüketim mallarının satıldığı ve perakende yapıda çok şubeli bir marketler zincirine ait veriler kullanılmıştır. Veriler, firmanın örnek bir şubeden alınarak sunulmuştur. Şubeden alınan veriler, 1 Ocak 2018 - 31 Mart 2018 tarihleri arasında, firmanın ciro olarak en fazla satış yaptığı ilk 10 ürün grubunda yapılan tüm alışverişleri arasından alınmıştır. Firmanın bir önceki yıl cirosunda satış hacmi en çok olan ilk 10 kategorisinin tüm cirodaki payı %26 olarak tespit edilmiştir. Kategori bazında cirodaki payları Tablo 1 sunulmuştur.

Tablo 1: İlk 10 Kategorinin Yüzdeler Payları

Kategoriler	İlk 10 kategori içindeki payı (%)	Tüm ciro içindeki payı (%)
deterjan	%14,58	%3,79
çay	%14,53	%3,77
kola	%14,41	%3,74
meyve suyu	%10,38	%2,69
su	%10,07	%2,62
bisküvi	%9,06	%2,35
soda	%8,50	%2,21
bebek bezi	%7,39	%1,92
şampuan	%5,99	%1,56
makarna	%5,09	%1,32
toplam	%100,00	%26,00

Örnek olarak incelenen şubeden ilgili dönemde 3409 adet müşteri (tekil) alışveriş yapmış ve bu müşteriler 11323 alışveriş gerçekleştirmişlerdir.

4 2 | Verinin Hazırlanması

İlgili tarihler arasında, ilgili ürün kategorilerinin satın alımlarından oluşmuş sepetler matris forma dönüştürülmüştür. Çalışmada, birliktelik kurallarının ortaya çıkarımı için ücretsiz ve açık kaynak kodlu bir veri madenciliği yazılımı olan WEKA kullanılmıştır. Şekil 2 'de örnek bir veri seti gösterilmektedir.

Şekil 2: Değişkenlerin Tanıtılması

relation veriler
@attribute DETERJAN numeric
@attribute ÇAY numeric
@attribute KOLA numeric
@attribute MEYVESUYU numeric
@attribute SU numeric
@attribute BİSKÜVİ numeric
@attribute SODA numeric
@attribute BEBEKBEZİ numeric
@attribute ŞAMPUAN numeric
@attribute MAKARNA numeric
@data
?, ?, ?, ?, 1, ?, ?, ?, ?
?, ?, ?, ?, 1, ?, ?, ?, ?
?, ?, ?, ?, 1, ?, ?, ?, ?
?, 1, ?, ?, 1, ?, ?, ?, ?
1, 1, ?, ?, ?, ?, 1, 1
1, ?, ?, ?, ?, ?, ?, ?

Şekil 2’ de görüldüğü gibi, değişkenler nümerik olarak tanıtılmıştır. Ancak bu nümeriklik tanımlanan değer ile alakalıdır. Veri kısmında, “1” kodlanması o ürünün alındığını, “?” ise alınmadığını ifade etmektedir. WEKA programında Apriori algoritması için parametrelerin seçildiği işlem penceresi Şekil 3’de gösterildiği gibidir.

Şekil 3: Apriori Algoritması Parametre Belirleme Ekranı



Şekil 3’ de görülen ekrandan minimum destek değeri, minimum güven değeri, ilginçlik ölçüsü (güven değeri, lift değeri, leverage değeri, conviction değeri), istenen sayıda kural vb. araştırmacının belirleyeceği parametreleri seçmek mümkündür.

Şekil 3’ de görülen ekranda, “numRules” sekmesi istenen maksimum kural sayısının yazılacağı kısımdır. “car” sekmesi, verideki herhangi bir değişkenin sınıf değişkeni olarak atanıp atanmayacağını belirlediği kısımdır. Burada “False” seçilirse sınıf değişkeni tanımlanmamış demektir. Bunun anlamı herhangi bir değişken hedef olarak belirlenmemiş yani kurallar iki yönlü olarak

elde edilebilir demektir. “delta” sekmesi minimum desteğe ulaşılan ya da gereken sayıda (numRules’daki sayı) kural üretilinceye kadar sağlanması gereken destek değerini vermektedir. “lowerBoundMinSupport” minimum destek için verilen alt sınırı göstermektedir. Bu sekme, incelenen veriler için 3 farklı değer için hazırlanmıştır. Analiz, daha yüksek ya da düşük bir destek değeri için yapmak istenirse bu sekmeyi değiştirmelidir. “metricType” ilginçlik ölçüsünün belirleneceği sekmedir. “minMetric” ise, araştırmacının yaratmak istediği kuralların minimum değerlerinin belirlendiği sekmedir. Araştırmacının buraya gireceği değerler üzerindeki değerlere sahip kurallar en iyi kural olarak karşısına çıkacaktır.

Analiz için üç farklı model tanımlanmış ve tanımlanan modellere ilişkin değerler Tablo 2’de sunulmuştur.

Tablo 1: Modeller ve Modellere Ait Kriterler

Kriterler	Modeller		
	1	2	3
lowerBoundMinSupport	0,01	0,005	0,005
metricType	Confidence	Confidence	Lift
minMetric	0,5	0,5	1,0

4.3 | Bulgular

Örnek şubede incelenen 11.323 sepetten alınan veriler modellerde yer alan değerlere göre analize tabi tutulmuş ve belirlendiği gibi 100 adet kural elde edilmesi istenmiştir. Analizler sonrası elde edilen kurallar arasından ilk 10 kurala ve ayrıca MS Excel programı vasıtasıyla hazırlanan formüller ile elde edilen destek ve güven değerleri ile beraber 22 adet ilginçlik ölçü birimi değerleri çalışmanın ekler kısmında sunulmuştur. Tablo 3’te üç farklı model için ilk kural ve bu kurala ait hesaplanan ilginçlik ölçü birimleri yer almaktadır.

Tablo 2: Sonuçlar

	Model 1:1.Kural	Model 2:1.Kural	Model 3:1.Kural
Kural	Öncül: çay,su, şampuan Ardıl: makarna soda,	Öncül: çay, bisküvi, şampuan Ardıl: makarna	Öncül: çay, şampuan Ardıl: deterjan, makarna
dd	0,0011	0,007	0,0053
gd	0,8125	0,693	0,1813
laplace correction	0,5002	0,501	0,4954
bütünlük değeri	0,0048	0,0291	0,1493
lift	3,3848	2,8869	5,1057
leverage	0,8122	0,6906	0,1802
certanity factor	0,7533	0,596	0,1511
klogsen	-0,2125	-0,1822	-0,0223
cosine	0,0623	0,1419	0,1645
jaccard	0,0048	0,0287	0,0892
gini index	-0,3618	-0,3488	-0,0794
yule's Q	0,8647	0,7599	0,7467
yule's Y	1,6036	1,4782	2,1914
müşterek güç	0,8741	0,884	0,9703
dkk	0,0927	0,22	0,2008
infomation gain	-8667,6944	-892,2464	-2192,8303
least contradiction	0,0037	0,0162	-0,5249
örnek ve karşı-örnek oranı	870,7692	142,8861	184,2
leverage-2	0,0008	0,0046	0,0043
coherence	0,0048	0,0287	0,0892
specifity-2	0,7606	0,7622	0,9456
kulczynski	0,4086	0,361	0,1653

Model 1 için 1. kural incelendiğinde; öncül kısmında 4 öge (çay, su, soda ve şampuan), ardıl kısmında 1 öge (Makarna) bulunmaktadır. Çay, su, soda ve şampuan kategorilerinden yapılan 16 satın alınanın 13 tanesinde, müşteriler aynı zamanda Makarna kategorisinden de 1 ürün almışlardır. Bir başka deyişle, çay, su, soda ve şampuan kategorisinden 1 ürün almış bir müşterinin %81,25 olasılık ile Makarna kategorisinden de 1 ürün alması beklenir (Ek 1).

Model 2'de belirlenen değerlere göre 100 kural oluşması istenmiş fakat 5 kural çıkmıştır. 1. kural incelendiğinde; öncül kısmında 3 öge (çay, bisküvi ve şampuan), ardıl kısmında 1 öge (makarna) bulunmaktadır. Çay, bisküvi ve şampuan kategorilerinden yapılan 114 satın alınanın 79 tanesinde, müşteriler aynı zamanda makarna kategorisinden de 1 ürün almışlardır. Bir başka deyişle, çay, bisküvi ve şampuan kategorisinden 1 ürün almış bir müşterinin %69,30 olasılık ile makarna kategorisinden de 1 ürün alması beklenir (Ek 2).

Model 3'ten elde edilen 1. kural incelendiğinde; öncül kısmında 2 öge (çay ve şampuan), ardıl kısmında 2 öge (deterjan ve makarna) bulunmaktadır. Çay ve şampuan kategorilerinden yapılan 331 satın alınanın 60 tanesinde, müşteriler aynı zamanda deterjan ve makarna kategorisinden de 1 ürün almışlardır. Bir başka deyişle, çay, bisküvi ve şampuan kategorisinden 1 ürün almış bir müşterinin %18,13 olasılık ile deterjan ve makarna kategorisinden de 1 ürün alması beklenir (Ek 3).

5 | SONUÇ ve DEĞERLENDİRME

Şirketlerin gelişen ve değişen rekabet koşullarında ayakta kalabilmelerinin yolu karar noktalarında bilimsel yöntem ve tekniklerden yararlanmalarına bağlıdır. Rekabette öne çıkmanın, fark yaratmanın en hayati olduğu konulardan biri de müşteri memnuniyetini sağlamak ve onların beklenti ve ihtiyaçlarını doğru belirleyebilmektir. Mevcut müşterilerini koruma, onların sadakatini kazanma ve potansiyel müşterileri kazanmanın yolu, müşteri beğeni ve satın alma alışkanlıklarını iyi ve doğru anlamak ile mümkündür. Buradan elde edilecek bilgiler işletmenin daha iyi ve uygun tasarlanmış bir mağaza yapısı ve tasarımına sahip olmasına yardımcı olacaktır. Bu da müşteri odaklı bir işletmeye giden yolda önemli bir adımdır. Müşteri beğeni ve alışkanlıkları, satın alma hareketleri izlenerek belirlenebilir. Bunun için de bu verileri sağlayabilecek metodlar kullanmak şarttır. Elde edilecek verilerden yola çıkarak, hangi ürün hangi ürün ile birlikte satılmakta, bu ürünlerin birbirleriyle beraber satılmaları arasında bir ilginçlik var mı gibi soruların yanıtına ulaşmak, şirketler için temel stratejiler geliştirmeleri açısından önemli olabilir. Şirketlerin tasarlayacakları kampanyalarda hangi grup müşteriye, hangi ürün gruplarıyla yönelmeli, dönemsel ve ya mevsimsel değişimler sırasında hangi ürün grupları ön plana çıkarılmalı, en az satan ürünlerin başka ürünlerle bir araya getirilmesinin nasıl etkileri olacağı konuları bu sayede belirlenebilir. Ayrıca mağaza yerleşimi, raf dizilimi, kasa önüne koyulabilecek özendirme malzemeleri de verilerden ortaya çıkarılabilecek sonuçlar neticesinde belirlenebilir. Bütün bunlar için veri madenciliği modellerinden biri olan birliktelik kuralları çıkarımı kullanılabilir. Bu kuralların ilginçlik ölçümleri ile değerlendirilmesi de gerekmektedir. Bu çalışma ana amaç olarak farklı ilginçlik ölçümlerini tanıtmayı belirlemiştir. Bunun yanında örnek bir şirketin bir şubesinde alışveriş verilerinde birliktelik kuralları analizleri gerçekleştirilmiştir. Elde edilen bulgular şirkete bir rapor olarak sunulmuştur. İzleyen kısımda bu öneriler özetlenmiştir.

İlginçlik olarak tespit edilmiş birliktelik kuralları için promosyon, ürün yerleştirme, e-posta pazarlaması veya kısa mesaj kampanyası yapılabilir. Kuraldaki bu

kategorilerden aynı anda alışveriş yapmamış müşterileri, bir başka deyişle böyle bir alışveriş alışkanlığı olmayan müşterileri, bu alışverişe teşvik etmek söz konusu olabilir. Doğru ürün ve doğru fiyat bir araya getirilmeli, marketin içinde promosyon bildirimleri buna göre yapılmalıdır. Yan yana konulmasında hijyenik, etik, resmi prosedürler, vb. gibi konularda sakınca olmayan ürünler için yan yana teşhir uygulamaları yapılabilir. İlginçlik Değerleri yüksek çıkmış kurallar içinde yer alan ürünlerin yan yana görülmesi, bir müşterinin tercih etmeyeceği, aklında olmadığı ya da ihtiyacı olmadığı için almayacağı bir ürünü almasına yol açabilir. Yan yana raflara dizilmese dahi, birliktelik kuralları içinde yer alan ürünlerin birinin bulunduğu rafın önüne (örneğin ketçap rafı) bir diğer ürün mağaza teşhir paleti üzerine (örneğin makarna) konabilir. Kasa önlerinde çalışanların önerileri doğrultusunda müşterilere sunulan küçük ürünlerin, ya da bazı promosyonların, birliktelik kuralları analizleri neticesinde ortaya çıkan ürünler arasından seçilebilir. Alışverişini tamamlamış, ödemesini yapmaya hazırlanan bir müşteriye kasada birliktelik kuralları arasında çıkmış ürünlerden birinin sepetinde olması durumunda, aynı kural içinde yer alan başka bir ürünün teklif edilmesi müşterinin aklında olmayan bir ürünü satın almasını sağlayabilir. Firmaya promosyon ürünleri belirleme, bandajlı ürün tasarlama, A ürününü alana B ürünü ücretsiz olarak veya belli bir indirim ile verilecektir türünden kampanyaları sistemleştirebilecekleri iletilmiştir. Birliktelik kuralları arasında olan ürünler arasında kasada anında bir indirim tasarlanması müşteriye beklenmeyen bir satın almaya yönlendirebilmektedir. Kurallar arasında zayıf ilginçlik değerleri olan bir ürün grubunda yapılan alışverişte diğer kategoriden alışverişini arttırması amacıyla diğer kategori ürünlerinde indirim olacağı bildirilebilir. Ancak bu işlemi duyurmak için kısa mesaj, e-mail ya da mağaza içi anons yayını gibi araçları kullanmak doğru olacaktır.

Bunların yanı sıra çalışmada bir takım sınırlılıklar, ileride yapılacak çalışmalara öneriler de vardır. Çalışmada müşteriler, demografik, cinsiyet, yaş vb. bir ayrıma tabi tutulmamış, yalnızca alışverişler incelenmiştir. Sonraki çalışmalarda müşteriler demografik, cinsiyet, yaş vb. ayrımlara tabi tutulup başka bir araştırma yapılabilir. Ayrıca veriler belirli bir dönemi içermektedir. Sonraki çalışmalarda yılın farklı dönemleri ya da yılın tümüne ait veriler için bir başka çalışma yapılabilir. Kurallar ve kurallara ait ilginçlik ölçümleri ayrı ayrı hesaplanmıştır. Ancak her bir ilginçlik ölçümü detaylı olarak yorumlama getirilmemiştir. Sonraki çalışmalarda birliktelik kuralları probleminin yapısına uygun ilginçlik ölçümleri tespit edilerek daha detaylı analizler gerçekleştirilebilir. Firmanın cirosundaki ilk 10 ürün kategorisindeki satın alma tercihlerinden elde edilen

veriler üzerinden yapılan bu çalışmanın dışında, tüm kategoriler üzerinden ya da firmanın daha fazla kar elde etmek, daha fazla ön plana çıkarmak istediği ürün gruplarından elde edilecek veriler üzerinden bir başka vaka analizi çalışması da gerçekleştirilebilir.

KAYNAKLAR

- Albarran, A. B., (2013). *The Social Media Instruments*, Routledge: New York
- Aras, R., Imirzi, H. O., & Akin, H. (2007). İstanbul'daki Küçük ve Orta Ölçekli Mobilya Üretim İşletmelerinin Sorunları ve Çözüm Önerileri, *Politeknik Dergisi*, 10(7): 105-110.
- Derham, R., Cragg, P., & Morrish, S. (2011). *Creating Value: An SME and Social Media*, Research in Agrawal, R., Imielinski, T. ve Swami, A. (1993). *Mining Association Rules between Sets of Items in Large Databases*. ACM Sigmod Conference, Washington DC.
- Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29 (1), 1-22.
- Alpaydın, E. (2000). *Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri*. Bilişim 2000 Eğitim Semineri.
- Aze, J. ve Kodratoff, Y. (2002). A Study of the Effect of Noisy Data in Rule Extraction Systems, *Proceedings of the Sixteenth European Meeting on Cybernetics and Systems Research, EMCSR'02*, 781-788.
- Azevedo, P. J. ve Jorge, A. M. (2007). Comparing Rule Measures for Predictive Association Rules, Editörler: J.N, Kok, J. Koronacki, R.L. Mantaras, S. Matwin, D. Mladenič, A. Skowron, *Machine Learning: ECML 2007. Lecture Notes in Computer Science*, Cilt 4701, Springer, Berlin, Heidelberg.
- Berzal, F., Blanco, I., Sanchez, D. ve Vila, M. A. (2002). Measuring the Accuracy and Interest of Association Rules: A New Framework”, *Intelligent Data Analysis*, 6, 221-235,
- Bramer, M. (2007). *Principles of Data Mining*, Springer-Verlag, London.
- Brin, S., Motwani, R., Ullman, J. ve Tsur, S. (1997). *Dynamic Itemset Counting and Implication Rules for Market Basket Data*, SIGMOD-97, 255-264.
- Dunham, M. H. (2006). *Data Mining: Introductory and Advanced Topics*, Prentice Hall, Upper Saddle River NJ.
- Fayyad, U., Piatetsky-Shapiro, G. ve Smyth, P. (1996). *From Data Mining to Knowledge Discovery: An Overview*. chapter 1: 1-34.
- Geng, L. ve Hamilton, H. J. (2006). Interestingness Measures for Data Mining: A Survey, *ACM Computing Surveys*, 38(3). Article: 9, 1-32, University of Regina,
- Gras, R., Briand, H., Peter, P. ve Philippe, J. (1996). Implicative Statistical Analysis, Proceedings of the Fifth Conference of the International Federation of Classification Societies, IFCS'96, 412-419.
- Jimenez, A., Berzal, F. ve Cubero, J. C. (2013). Interestingness Measures for Association Rules Within Groups, *Intelligent Data Analysis*, 17, 195–215.
- Kannan, S. ve Bhaskaran, R. (2009). Association Rule Pruning Based On Interestingness Measures with Clustering, *International Journal of Computer Science Issues (IJCSI)*, (6)1, 35-43.
- Kirchgesner, M., Leroy, V. ve Amer-Yahia, S. (2016). *Testing Interestingness Measures in Practice: A Large-Scale Analysis of Buying Patterns*, Shashwat Mishra University, France.
- Larose, D.T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc., New Jersey.
- Ledolter, J. (2013). *Data Mining and Business Analytics With R*. John Wiley & Sons, Inc., New Jersey.
- Lee, S. J. ve Siau, K. (2001). A Review of Data Mining Techniques, *Industrial Management & Data Systems*, 101(1), 41-46.
- Masood, A. ve Ouaguenouni, S. (2013). Probabilistic Measures for Interestingness of Deviations – A Survey”, *International Journal of Artificial Intelligence & Applications (IJAIA)*, 4(2), 39-61.
- McGarry, K. (2005). A Survey of Interestingness Measures for Knowledge Discovery, *The Knowledge Engineering Review*, Cambridge University Press, 1-24.
- Moen, P. (2005). *Data Mining Methods*. Lecture slides.
- Piatetsky-Shapiro, G. (1991). *Discovery, Analysis and Presentation of Strong Rules*. Knowledge Discovery in Databases, Editörler: G. Piatetsky-Shapiro, W. J. Frawley, AAAI MIT Press, 229-248.
- Rijsbergen, C. J. (1979). *Information Retrieval*. 2nd Edition. London: Butterworths,
- Sebag, M. ve Schoenauer, M. (1998). Generation of Rules with Certainty and Confidence Factors from Incomplete and Incoherent Learning Bases”, *EKA'88, Proceedings of the European Knowledge Acquisition Workshop*. Gesellschaft für Mathematik und Datenverarbeitung mbH, 1-20.
- Shaikh, M., McNicholas, P., Antonie, M. L. ve Murphy, T. B. (2013). Standardizing Interestingness Measures for Association Rules, *Statistical Analysis and Data Mining*, 11 (6), 282-295.
- Silahtaroglu, G. (2008). *Veri Madenciliği*, Papatya Bilim, İstanbul,
- Smyth, R. ve Goodman, R. M. (1991). Rule Induction Using Information Theory, Knowledge Discovery in Databases, Editörler: G. Piatetsky-Shapiro, W. J. Frawley, MIT Press, 159-176.

- Tan, P. N. ve Kumar, V. (2000). *Interestingness Measures for Association Patterns: A Perspective*, Department of Computer Science and Engineering University of Minnesota.
- Tan, P. N., Kumar, V. ve Srivastava, J. (2004). Selecting the Right Objective Measure for Association Analysis, *Information Systems*, 29, 293–313.
- Tan, P. N., Steinbach, M. ve Kumar, V. (2006). *Introduction to Data Mining*, Pearson Addison-Wesley.
- University of Helsinki, Department of Computer Science.
<http://www.cs.helsinki.fi/u/ronkaine/tilome/luentomateriaali/TiLoMe-170105.pdf>, (Erişim Tarihi: 25.11.2018)
- Zaki, M. J. ve Meira, W. Jr. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press, Cambridge.
- Zhao, Y., Zhang, C. ve Cao, L. (2009). *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*, New York: Information Science Reference, Hershey.

Ek 1: Örnek Şube - 1. Kritere Göre Oluşan İlk 10 Kural

	dd	gd	laplace correction	bütünlük değeri	lift	leverage	certaninity factor	klosgen	cosine	jaccard	gini index	yule's Q	yule's Y	collective strength	doğrusal korelasyon katsayısı	infomation gain	last contradiction	Örnek ve karşı örnek oranı	leverage-2	coherence	specifity-2	kulczyński
çay, su, soda, şampuan => makarna	0,001 1	0,812 5	0,500 2	0,004 8	3,384 8	0,812 2	0,753 3	- 0,212 5	0,062 3	0,004 8	- 0,361 8	0,864 7	1,603 6	0,874 1	0,092 7	- 8667,6 9	0,003 7	870,7 692	0,000 8	0,004 8	0,760 6	0,408
deterjan, çay, bebekbezi => makarna	0,001 9	0,758 6	0,500 3	0,008 1	3,160 4	0,758 0	0,682 4	- 0,206 6	0,078 4	0,008 1	- 0,360 0	0,818 6	1,549 1	0,875 5	0,118 4	- 4410,5 3	0,005 5	514,3 636	0,001 3	0,008 1	0,760 8	0,383
deterjan, çay, bisküvi, şampuan38 => makarna	0,002 4	0,710 5	0,500 4	0,009 9	2,960 0	0,709 7	0,619 1	- 0,205 3	0,084 0	0,009 9	- 0,359 2	0,773 7	1,498 9	0,876 0	0,129 1	- 3255,5 2	0,005 9	418,9 630	0,001 6	0,009 9	0,760 8	0,360
çay, bisküvi, şampuan => makarna,	0,007 0	0,693 0	0,501 0	0,029 1	2,886 9	0,690 6	0,596 0	- 0,182 2	0,141 9	0,028 7	- 0,348 8	0,759 9	1,478 2	0,884 0	0,220 0	- 892,24	0,016 2	142,8 861	0,004 6	0,028 7	0,762 2	0,361
deterjan, çay, meyvesuyu, makarna=> bisküvi	0,001 7	0,678 6	0,500 2	0,006 0	2,413 1	0,677 9	0,552 8	- 0,253 4	0,063 6	0,006 0	- 0,400 6	0,688 6	1,316 3	0,850 6	0,106 2	- 3991,0 5	0,003 1	595,4 737	0,001 0	0,006 0	0,719 2	0,342
çay, meyvesuyu, bisküvi, soda, şampuan=> makarna	0,001 0	0,647 1	0,500 1	0,004 0	2,695 6	0,646 7	0,535 6	- 0,219 9	0,051 2	0,004 0	- 0,362 7	0,706 9	1,430 8	0,873 3	0,080 5	- 8359,1 4	0,001 8	1028, 8182	0,000 6	0,004 0	0,760 2	0,325
çay, bisküvi, soda, şampuan=> makarna	0,002 0	0,638 9	0,500 2	0,008 5	2,661 6	0,638 1	0,524 8	- 0,211 2	0,073 5	0,008 4	- 0,360 5	0,698 8	1,421 1	0,875 0	0,116 2	- 3527,6 3	0,003 7	491,7 391	0,001 3	0,008 4	0,760 4	0,323
deterjan, meyvesuyu, bebekbezi=> makarna	0,001 2	0,636 4	0,500 1	0,005 2	2,651 0	0,635 9	0,521 5	- 0,217 7	0,057 3	0,005 1	- 0,362 2	0,695 3	1,418 7	0,873 7	0,090 5	- 6234,7 6	0,002 2	808,2 143	0,000 8	0,005 1	0,760 2	0,320
meyvesuyu, su, soda, şampuan=> makarna	0,001 2	0,636 4	0,500 1	0,005 2	2,651 0	0,635 9	0,521 5	- 0,217 7	0,057 3	0,005 1	- 0,362 2	0,695 3	1,418 7	0,873 7	0,090 5	- 6234,7 6	0,002 2	808,2 143	0,000 8	0,005 1	0,760 2	0,320
deterjan, çay, bisküvi => makarna	0,004 9	0,629 2	0,500 5	0,020 6	2,621 3	0,627 3	0,512 1	- 0,195 8	0,113 9	0,020 4	- 0,354 5	0,690 6	1,408 7	0,879 5	0,180 9	- 1222,0 8	0,008 5	201,6 071	0,003 1	0,020 4	0,760 8	0,324

Ek 2. Örnek Şube - 2. Kritere Göre Oluşan İlk 5 Kural

	dd	gd	laplace correction	bütünlük değeri	lift	leverage	certainty factor	kloggen	cosine	jaccard	gini index	yule's Q	yule's Y	collective strength	doğrusal korelasyon katsayısı	infomation gain	last contradiction	Örnek ve karşı örnek oranı	leverage-2	coherence	specifity-2	kulczynski
çay, bisküvi, şampuan=> makarna	0,0070	0,6930	0,5010	0,0291	2,8869	0,6906	0,5960	-0,1822	0,1419	0,0287	-0,3488	0,7599	1,4782	0,8840	0,2200	-892,2464	0,0162	142,8861	0,0046	0,0287	0,7622	0,3610
deterjan, çay, şampuan=> makarna	0,0053	0,5941	0,5004	0,0221	2,4748	0,5919	0,4658	-0,1968	0,1145	0,0217	-0,3544	0,6500	1,3681	0,8794	0,1849	-1062,8898	0,0070	188,0333	0,0032	0,0217	0,7604	0,3081
meyvesuyu, bisküvi, şampuan=> makarna	0,0065	0,5564	0,5003	0,0272	2,3179	0,5536	0,4163	-0,1951	0,1231	0,0266	-0,3529	0,6043	1,3227	0,8802	0,2028	-774,8536	0,0055	152,2162	0,0037	0,0266	0,7597	0,2918
çay, şampuan=> makarna=1	0,0069	0,2356	0,4962	0,0287	0,9817	0,2286	-0,0058	-0,2205	0,0822	0,0263	-0,3658	-0,0125	0,8509	0,8589	0,1911	-308,0878	-0,0644	141,9231	-0,0001	0,0263	0,7427	0,1322
bisküvi, şampuan => makarna	0,0187	0,5000	0,5000	0,0780	2,0830	0,4910	0,3421	-0,1716	0,1975	0,0724	-0,3358	0,5401	1,2426	0,8911	0,3385	-192,2018	0,0000	52,4104	0,0097	0,0724	0,7554	0,2890

Ek 3. Örnek Şube 3. Kritere Göre Oluşan İlk 10 Kural

	dd	gd	laplace correction	bütünlük değeri	lift	leverage	certainty factor	kloggen	cosine	jaccard	gini index	yule's Q	yule's Y	collective strength	doğrusal korelasyon katsayısı	infomation gain	last contradiction	Örnek ve karşı örnek oranı	leverage-2	coherence	specifity-2	kulczynski
Çay, şampuan => deterjan, makarna	0,0053	0,1813	0,4954	0,1493	5,1057	0,1802	0,1511	-0,0223	0,1645	0,0892	-0,0794	0,7467	2,1914	0,9703	0,2008	-2192,8303	-0,5249	184,2000	0,0043	0,0892	0,9456	0,1653
Deterjan, makarna=> çay, şampuan	0,0053	0,1493	0,4939	0,1813	5,1057	0,1482	0,1236	-0,0184	0,1645	0,0892	-0,0743	0,7467	2,1914	0,9671	0,2002	-2192,8303	-0,8520	183,0167	0,0043	0,0892	0,9456	0,1653
Deterjan, çay => şampuan, makarna	0,0053	0,2120	0,4964	0,1105	4,4211	0,2108	0,1723	-0,0325	0,1531	0,0783	-0,0977	0,7093	2,0303	0,9680	0,1929	-1898,7719	-0,3002	185,0000	0,0041	0,0783	0,9373	0,1613
Şampuan, makarna => deterjan, çay	0,0053	0,1105	0,4909	0,2120	4,4211	0,1093	0,0877	-0,0169	0,1531	0,0783	-0,0789	0,7093	2,0303	0,9566	0,1906	-1898,7719	-1,4947	180,6667	0,0041	0,0783	0,9373	0,1613
Deterjan, şampuan => çay, makarna	0,0053	0,2076	0,4963	0,0952	3,7314	0,2062	0,1609	-0,0405	0,1406	0,0698	-0,1118	0,6558	1,8570	0,9633	0,1840	-1602,5835	-0,2683	184,9000	0,0039	0,0698	0,9291	0,1514
Çay, makarna => deterjan, şampuan	0,0053	0,0952	0,4890	0,2076	3,7314	0,0938	0,0715	-0,0186	0,1406	0,0698	-0,0875	0,6558	1,8570	0,9483	0,1811	-1602,5835	-1,7647	179,2167	0,0039	0,0698	0,9291	0,1514
Çay, şampuan => bisküvi, makarna	0,0070	0,2387	0,4962	0,1045	3,5747	0,2367	0,1842	-0,0468	0,1579	0,0784	-0,1290	0,6538	1,8046	0,9592	0,2099	-1104,8144	-0,2288	140,1392	0,0050	0,0784	0,9174	0,1716
Bisküvi, makarna=> çay, şampuan	0,0070	0,1045	0,4872	0,2387	3,5747	0,1025	0,0775	-0,0205	0,1579	0,0784	-0,1001	0,6538	1,8046	0,9405	0,2058	-1104,8144	-1,8066	134,7595	0,0050	0,0784	0,9174	0,1716
Meyvesuyu, şampuan => bisküvi, makarna	0,0065	0,2242	0,4960	0,0979	3,3586	0,2223	0,1687	-0,0486	0,1482	0,0731	-0,1305	0,6276	1,7488	0,9577	0,1997	-1122,7550	-0,2407	149,5541	0,0046	0,0731	0,9166	0,1611
Bisküvi, makarna => meyvesuyu, şampuan	0,0065	0,0979	0,4870	0,2242	3,3586	0,0959	0,0708	-0,0212	0,1482	0,0731	-0,1012	0,6276	1,7488	0,9390	0,1958	-1122,7550	-1,8424	143,7973	0,0046	0,0731	0,9166	0,1611