

VERİ MADENCİLİĞİNDE SEPET ANALİZİ İLE TÜKETİCİ DAVRANIŞI MODELLEMESİ

Doç. Dr. Mehpere TİMOR

Dr. Umman Tuğba ŞİMŞEK

İstanbul Üniversitesi

İşletme Fakültesi

Sayısal Yöntemler Anabilim Dalı

Veri madenciliği büyük miktardaki verinin, bu verilerden anlamlı örüntü ve kurallar çıkarılması amacıyla analiz edilmesidir. Veri madenciliğinde verinin yapısına uygun olarak birçok teknik kullanılmaktadır. Bu tekniklerden biri olan “Birliktelik Kuralları Analizi/Sepet Analizi” ile, büyük veri tabanlarından yola çıkılarak, müşterilerin alışveriş davranışları keşfedilmeye çalışılmaktadır. Bu çalışmada, Türkiye’de perakende sektöründe faaliyet gösteren büyük bir market zincirine ait veriler ele alınarak, veriler “Birliktelik Kuralları” ile analiz edilmiş ve müşterilerin satın alma davranışlarını etkileyen değişkenler karar ağaçlarıyla belirlenmiştir. Çalışmada verilerin analizi için SPSS’in veri madenciliği modülü olan Clementine programı kullanılmıştır.

Anahtar Sözcükler: Veri Madenciliği, Birliktelik Kuralları Analizi, Sepet Analizi, Karar Ağaçları.

CUSTOMER BEHAVIOR MODELING BY USING MARKET BASKET ANALYSIS IN DATA MINING

Data mining analyzes large amount of data for the purpose of deriving meaningful patterns and rules. Many techniques in accordance with the structure of data are used in data mining. with “Association Rules Analysis/Market Basket Analysis”, that is one of these techniques; shopping behaviors of customers are tried to be found out by setting out from large data bases. In this study, data belonging to a big market chain that activates in the retail sector in Turkey was handled, the data was analysed with “Association Rules” and the variables effecting purchasing behaviours of costumers were determined with decision trees. In the study, Clementine program, the data mining module of SPSS, was used for analysis of the data.

Key Words: Data Mining, Association Rules Analysis, Market Basket Analysis, Decision Trees.

GİRİŞ

Günümüz dünyasında işletmeler başarılı olabilmek için yeni ekonominin ve yeni iş modellerinin kurallarına uygun hareket etmek durumundadır. Bilgisayar ve iletişim teknolojilerindeki gelişmeler çok daha fazla verinin hızlı toplanmasına, depolanmasına, işlenmesine ve bilgiye dönüştürülüp yeniden istenilen noktalara iletilmesine olanak sağlamıştır. Yeni teknolojilerden önemli ölçüde etkilenen bu yeni ekonominin kuralları gereği “bilgi” ve “zaman” boyutlarının önemi daha çok artmış ve dolayısıyla işletmelerin karar verme süreçlerinde doğru ve anlamlı bilgiye dayalı hızlı karar alma gereği her zamankinden daha fazla ön plana çıkmıştır. Saklanması gereken verilerin miktar ve çeşitliliğinin artması, analizlerin daha hızlı yapılması gereği, sonuçta anlamlı ve eyleme yönelik bilgiler ortaya çıkarılması, yeni ekonominin ve değişen piyasalardaki yoğun rekabet ortamının bir gereğidir. Milyonlarca bitlik verilerin depolandığı veri ambarlarının tutulduğu donanımlar ve bunlardan çok hızlı bir şekilde anlamlı bilgiler elde edebilmek için kullanılabilecek yazılımlar yeni teknolojinin sunduğu olanaklardır. (Özmen, 2001)

Çoğu zaman iyi kullanılmamaları durumunda veri tabanlarında tutulan veri, problem haline gelebilmektedir. Toplanan veri miktarı büyüdükçe ve toplanan verilerdeki karmaşıklık arttıkça, daha iyi çözümlene tekniklerine olan gereksinim de artmaktadır. Bu noktada “veri madenciliği” ve “veri tabanlarında bilgi keşfi” kavramları ortaya çıkmaktadır. Eğer veri tabanlarında bilgi keşfi süreci başarılı ise keşfedilen bilgi, organizasyonların karar verme sürecini geliştirmek amacıyla kullanılabilir. (Freitas, 2002, s.2)

Veriden anlamlı bilgilerin çıkarılması sürecine “bilgi keşfi” denmektedir. Veri madenciliği ise bu süreç içerisinde yer alan önemli bir adımdır. Veri madenciliği, spesifik algoritmaların, veride bulunan anlamlı örüntüleri belirlemek için uygulanması sürecidir. (Fayyad, Piatetsky-Shapiro ve Smyth, 1996)

Veri Madenciliğindeki amaç, toplanmış olan bilgilerin, bir takım sayısal yöntemlerle incelenmesi ve çok büyük veri tabanlarındaki ya da veri ambarlarındaki veriler arasında bulunan ilişkiler, örüntüler, değişiklikler, belirli yapılar gibi ilginç bilgilerin ortaya çıkarılması ve keşfi işlemidir. Bu çalışmada öncelikle veri madenciliği tekniklerinden “Birliktelik Kuralları” kullanılarak müşterilerin alışveriş davranışları belirlenmiştir. Birliktelik Kuralları analizini gerçekleştirmek için veriler yeniden kodlanarak gerekli düzenlemeler yapılmıştır. Gerçekleştirilen analizde, müşterilerin en fazla satın aldığı ürünler hedef değişken olarak tanımlanmış ve “Karar Ağaçları” kullanılarak müşterilerin satın alma davranışını en çok etkileyen değişkenler belirlenmiştir.

1. LİTERATÜR TARAMASI

Veri madenciliği (VM) yaygın bir uygulama alanına sahiptir. Literatürde VM kullanılarak müşteri davranışlarının incelendiği pek çok çalışma mevcuttur. Chen (2007) çalışmasında veri zarflama analizi kullanarak, birliktelik kurallarının etkinliğini çok amaçlı karar verme yaklaşımıyla belirlemek istemiştir. Birliktelik kuralları destek ve güven kriterleri ile ölçülmektedir.

Chen ve diğerleri (2005) mağazaların alışveriş kayıtlarının tutulduğu veri tabanlarını kullanarak müşterilerin satın alma davranışlarını belirleyebilmek için bir sepet analizi uygulaması yapmışlardır. Analiz sonucunda elde edilen bilgiler pazarlama, satış ve operasyon stratejilerinin şekillendirilmesinde önem taşımaktadır. Varolan diğer teknikler birden çok mağazanın bulunduğu zincirlerde, mağaza sayısının artmasına ve ürün yelpazesinin genişlemesine bağlı olarak önemli satınalma örüntülerinin belirlenmesinde eksik kalabilmektedir. Chen ve diğerleri çalışmalarında kullandıkları teknikle, geleneksel tekniklerin aksine bu zayıflığın üstesinden gelmek istemişlerdir.

Vindevoel, Van den Poel ve Wets'in (2005) çalışmalarında büyük miktardaki ürün çiftleri arasındaki bağıntıyı hesaplayabilmek için sepet analizi kullanılmış, promosyon stratejilerini optimize etmek için bir öneri getirilmiştir. Birlikte satın alınan ürünlerden sadece bir tanesinin fiyatının düşürülmesi ile etkin sonuçlar yaratacağı ileri sürülmüştür. Promosyonların kısa vadeli ve uzun vadeli kalıcı etkileri, çok değişkenli zaman serileri analizi kullanılarak analiz edilmiştir.

Veri madenciliği uygulamalarında birliktelik kurallarını tespit edebilmek oldukça önemlidir. Veri tabanlarında gizlenmiş olan ilişkilerin tespit edilmesi ile elde edilecek bilgiler sepet analizi veya müşteri profilini belirleme gibi pek çok amaç için kullanılabilir. Angiulli, Ianni ve Palopoli (2004) çalışmalarında birliktelik kurallarını sayısal algoritmalarla belirlemeye çalışmışlardır.

Perakende mağaza yöneticileri uzun yıllardır müşterilerinin satınalma davranışlarını incelemektedir. Son zamanlarda perakende mağazalarda çapraz satış örüntülerini belirlemek önem taşımaya başlamıştır. Mild ve Reutterer (2003) yaptıkları çalışmada müşteri tercihlerini veya önceliklerini belirlemek için collaborative filtering (ortak filtreleme-işbirliğine yönelik süzgeç) algoritmasını kullanmıştır. Yaptıkları çalışmada bu teknik incelenmiş, çapraz satış örüntüleri belirlenmeye çalışılmış ve bir perakende mağazanın veri tabanında uygulama yapılmıştır.

Giudici ve Passerone (2002), müşteri davranışları ile ilgili olarak grafik yöntemi sepet analizine uyarlayarak bir çalışma yapmıştır. “Sepet analizi” müşterilerin, örneğin bir süpermarketteki farklı ürün

alışverişleri ile ilgili olarak birliktelik kurallarının araştırmaya yöneliktir. Birliktelik kurallarının tespit edilmesi, pazarlama planlarının oluşturulmasında yarar sağlamaktadır. Eğer iki ürün arasında yüksek oranda bir bağıntı varsa, mağaza raf düzenlemelerinde bu iki ürünü yan yana koymak satışları arttıracaktır.

Hruschka, Lukanowicz ve Buchta (1999), çalışmalarında çok boyutlu Binomial Logit modelini kullanmışlardır. Bu model, müşterilerin pazar sepet verilerini ve mağazada promosyonda olan ürün kategorileri bilgilerini gerektirmektedir. Çalışmada modelin parametrelerinin belirlenme süreci adım adım anlatılmıştır. Orta büyüklükteki bir süpermarkete ait 6147 satınalma hareketi incelenmiş ve gerçekleştirilen uygulama sonucunda bu modelin perakendeci firmanın promosyon kararları üzerindeki etkisi incelenmiştir.

Yöneylem Araştırması ve Endüstri Mühendisliği Doktora öğrencileri (2003), müşteri tabanında saklanan her bir müşterinin firma içinde veya firma ile gerçekleştirdiği işlemlere ait kayıtlarını kullanarak, kişi bazında tüm müşterilerin firmaya bağlılığını ve firma için değerini belirlemek amacıyla bir çalışma gerçekleştirmiştir. Çalışmada firma ile iş yapmış tüm müşterilere ve bu müşterilerin yaptıkları işlemlere ait zaman ve işlem sıklığı bilgileri kullanılarak, şu sorulara cevaplar aranmıştır: Müşteri tabanındaki kişilerden hangileri “aktif”, hangileri “pasif” müşteridir? Bu müşterilerin şirkete “bağlılık” seviyeleri nelerdir? Firmanın mevcut “aktif” müşteri tabanı bir önceki yılda ne kadar büyümüştür? Bu listedeki müşterilerin, gerek bireysel gerekse toplu olarak, bir sonraki yıl ne kadarlık ve ne tür işlemler yapması beklenmelidir?

Sever ve Oğuz (2002) çalışmalarında Biçimsel Kavram Analizi kullanarak eşleştirme kuralı çıkarım problemi için bir çerçeve geliştirmişlerdir. Çalışmada eşleştirme sorgularının özel bir durumu olan sepet analizi kullanılmıştır. Sepet analizi, bir firmada müşterilerin satınalma eğilimleri hakkında bilgi verir. Sepet analizinde amaç, verilen minimum güvenilirlik ve destek kriterlerini sağlayan ilginç eşleştirmeleri bulmaktır.

2. VERİ MADENCİLİĞİ

Veri madenciliği pek çok alanda yaygın olarak kullanılan, Türkiye’de ve dünyada özellikle son yıllarda oldukça önem kazanmış olan bir tekniktir. Veri madenciliği; büyük miktarda veri içinden, gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanılarak aranmasıdır. Veri analizi yapılarak, bir mal için bir sonraki ayın satış tahminleri yapılabilir, müşteriler satın aldıkları mallara bağlı olarak gruplanabilir, yeni bir ürün için potansiyel müşteriler belirlenebilir, müşterilerin zaman içindeki

hareketleri incelenerek onların davranışları ile ilgili tahminler yapılabilir. Binlerce malın ve müşterinin olabileceği düşünülürse bu analizin gözle ve elle yapılamayacağı, otomatik olarak yapılmasının gerektiği ortaya çıkar ve veri madenciliği bu noktada devreye girer. (Alpaydın, 2000, s.1)

Veri madenciliği günümüzde perakende, bankacılık, telekomünikasyon ve daha birçok farklı alanda kullanılmaktadır. (Rygielski, Wang ve Yen, 2002) Perakende sektöründeki uygulamalara aşağıda örnek verilmiştir:

- **Sepet Analizi:** Müşterilerin hangi ürünleri hangi ürünlerle birlikte satın aldığı analiz edilir. Elde edilecek olan bilgi mağaza raf düzenlemelerinde ve promosyon stratejilerinin belirlenmesinde kullanılır.

- **Satış Tahminleri:** Perakendeciler satış tahminlerini stok kontrolünde kullanırlar. Eğer bir müşteri bugün alışveriş yaparsa, izleyen dönemde ne zaman alışveriş yapacaktır? sorusuna cevap aranır.

Veri madenciliği pazarlama alanında müşterinin fiyat artışı ile değişen satınalma alışkanlıklarının belirlenmesi, çapraz satış analizleri ile ürün satışları arasındaki birlikteliklerin ve ilişkilerin belirlenmesi, müşteri profili belirleme çalışmaları kapsamında hangi özelliklerdeki müşterilerin hangi ürünleri satın aldıklarının belirlenmesi amacıyla kullanılmaktadır. (www.spss.com.tr)

Veri madenciliği bankacılık sektöründe işletme riskinin azaltılması, risk derecelendirme tahmini, maliyetlerin düşürülmesi, karlılık analizleri, ATM’lere gün içinde dağıtılacak para miktarının tespiti, doğru ve etkin kredi kararı verebilme, müşteri ve çalışan memnuniyetinin artırılması için kullanılmaktadır. (Özel, 2003)

Telekomünikasyon sektöründe en önemli sorun müşteri kaybıdır. Kuruluşlar hangi müşterilerini kaybedebileceklerini önceden belirleyebildikleri takdirde bu müşterilerini elde tutma amaçlı stratejiler geliştirebilir, düşük maliyetli ve etkili kampanyalar düzenleyebilirler. Türkiye’de mobil iletişimin büyüme hızı ve pazardaki rekabet göz önüne alındığında yöneticilerin bilgi ve tecrübelerinin şirketin işleyişi esnasında tutulan verilerde gizli olan bilgi ile desteklenmesi ve buna göre stratejiler belirlenmesinin önemi ortadadır. (www.spss.com.tr)

Veri madenciliğinin farklı uygulama alanları için yapılmış olan çalışmalar da mevcuttur. Wang ve Weigend (2003) finans alanında veri madenciliği ile bir portföy yönetimi uygulaması gerçekleştirmişlerdir. Wei ve Chiu (2002) telekomünikasyon sektöründe müşteri kaybını tahmin edebilmek için veri madenciliğini kullanmışlardır. Min ve Emam (2002) turizm sektöründe otel müşterilerinin önceliklerinin belirlenmesi ve müşteri sadakatının artırılması için veri madenciliği tekniğini kullanmışlardır. Liu ve diğerleri

(2001) veri madenciliğini gıda sektöründe, zaman serileri verilerini içeren fast-food restaurantına uygulamışlardır.

3. VERİ MADENCİLİĞİNDE KULLANILAN MODELLER

Veri madenciliğinde kullanılan birçok model söz konusudur. Veri madenciliğinde kullanılan modeller; ilişki analizi, sınıflandırma, kümeleme ve tahmin edici modeller olmak üzere 4 başlık altında incelenmektedir. (Moss, 2003, s. 307)

Birliktelik kuralları ve ardışık zamanlı örüntüler, “İlişki Analizi” altında yer alır ve pazarlama amaçlı olarak *pazar sepet analizi (market basket analysis)* adı altında veri madenciliğinde yaygın olarak kullanılmaktadır.

Birliktelik kuralları veri madenciliğinin en iyi örneklerinden biridir. Birliktelik kuralları verideki potansiyel ilişkileri tanımlar. Birliktelik kuralları analizi ile büyük veri tabanlarından, müşterilerin alışveriş davranışları keşfedilmeye çalışılır. Birliktelik kuralları bir ürün satın alındığında, o ürünle beraber başka hangi ürünlerin satın alındığının belirlenmesi amacıyla kullanılır. (Yen ve Lee, 2006) Birliktelik kuralları, eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır. Örneğin; tatil dolayısıyla tüm aile bireylerine uçak bileti alan bir müşteri, % 95 olasılıkla tatil beldesinde araba da kiralayacaktır. Perakende mağazalar bu tekniği, müşterilerinin satınalma eğilimlerini belirlemek için kullanmaktadır. (Akpınar, 1998)

Birliktelik kurallarının kullanıldığı en tipik örnek “Sepet Analizi”dir. Sepet analizi ile, müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikler bulunarak, müşterilerin satın alma alışkanlıkları belirlenir. Bu tip birlikteliklerin belirlenmesi, müşterilerin hangi ürünleri bir arada aldıkları bilgisini ortaya çıkarır ve market yöneticileri de bu bilgi ışığında raf düzenlerini belirleyerek, ürün paketleri hazırlayarak satış oranlarını artırabilir ve etkili satış stratejileri geliştirebilirler. Sepet analizinde amaç alanlar arasındaki ilişkileri bulmaktır. Bu ilişkilerin bilinmesi şirketin karını arttırmak için kullanılabilir. Eğer X malını alanların Y malını da çok yüksek olasılıkla aldıkları biliniyorsa ve eğer bir müşteri X malını alıyor ama Y malını almıyorsa o müşteri potansiyel bir Y müşterisidir.

Birliktelik kurallarının belirlenmesinde birçok yöntem kullanılmaktadır. Birliktelik kurallarıyla ilgili problem, birliktelik kurallarını bulmada bir eşik değeri bulmaktır. Önemsiz gürültüden değerli bilgiyi ayırabilmek ve bu eşik değerini bulabilmek çok zordur. Bu yüzden ilginç birliktelik kurallarından ilginç olmayanları ayırt edebilmek için bazı ölçütlerin

belirlenmesi gereklidir. Bu ölçütler destek (support) ve güven (confidence) kriterleridir. (Dolgun ve Zor, 2006)

Elde olan veride ürünler için sadece satın alındı/alınmadı bilgisi varsa sepet analizinde ürünler arasındaki bağıntı, destek ve güven kriterleri aracılığı ile hesaplanır. Güçlü bir kural çok büyük bir destek ve yüksek seviyeli bir güvenilirliğe sahiptir. (Tuğ ve Bulun, 2006)

Destek: $P(X \text{ ve } Y) = X \text{ ve } Y \text{ mallarını satın almış müşteri sayısı} / \text{Toplam müşteri sayısı}$

(X ve Y ürünlerini içeren hareketlerin sayısı)

Güven: $P(X/Y) = P(X \text{ ve } Y) / P(Y)$
 $= X \text{ ve } Y \text{ mallarını satın almış müşteri sayısı} / Y \text{ malını satın almış müşteri sayısı}$

(X ürününün bulunduğu yerde X ve Y ürün setinin bulunma olasılığı)

Destek kriteri veride bu bağıntının ne kadar sık olduğunu, güven kriteri ise Y malını almış bir kişinin hangi olasılıkla X malını alacağını belirtmektedir. İki ürünün satın alınmasındaki bağıntının önemli olması için her iki değer de olabildiğince yüksek olması gerekmektedir. (Alpaydın, 2000, s.9)

4. BİRLİKTELİK KURALLARI VE SEPET ANALİZİ İLE TÜKETİCİ SATIN ALMA DAVRANIŞI MODELLEMESİ

Bu çalışmada kullanılan veriler Türkiye'nin en büyük market zincirlerinden birinin 2004 yılına ait Şubat, Mart, Nisan ve Mayıs ayları müşteri alışveriş kayıtlarından oluşmaktadır. Belirtilen dört aylık periyotta 999 tüzel olmayan gerçek müşteri üzerinde yapılan araştırmada, bu müşterilerin 628'ine ait 27.924 alışveriş kaydı kullanılmıştır. 60 farklı ürün kategorisi, 438 ürün grubu ve 6018 ürünü içeren bu uygulamada müşterilerin tercih ettikleri 971 farklı ürün yer almaktadır. 999 müşterinin 322'si evli kadınlardan, 173'ü bekar kadınlardan, 363'ü evli erkeklerden ve 141'i bekar erkeklerden oluşmaktadır.

Analizde SPSS Clementine programı kullanılmıştır. Program veri girişinin yapılmasına, farklı veri setlerini birleştirmeye, modellemeye ve analiz etmeye olanak sağlamaktadır. Sistem yapay sinir ağları ve kural çıkarımı temeline dayanmaktadır. Sonuçlar anlaşılması kolay grafiklerle ve görselleştirme teknikleriyle sunulmaktadır. Clementine analiz yapan kişilere, kısa sürede sonuca ulaşmayı sağlayacak çok geniş bir fonksiyon seti sunmaktadır.

Veri madenciliği ve bilgi keşfi konusunda yayın ve araştırma yapan KDnuggets dergisi, “En çok tercih edilen veri madenciliği çözümü” konulu bir araştırma yapmış, üçüncüsü düzenlenen uluslararası geleneksel anket çalışmalarının tümünde uzman sektörel kullanıcılar, kendilerine yöneltilen sorulara verdikleri

Tablo1- Veri Seti

Müşteri No	Kategori	Adet Tür	Adet	Birim Fiyat	Coğrafi Bölge	Şehir	Gerçek/ Tüzel	Cinsiyet	Medeni Hal	Yaş	Üyelik Yılı
128402	Şekerli Ürünler	1	2	390.000	Akdeniz	İçel	Gerçek	Erkek	Evlü	33	1
128402	Meşrubat ve Kuruyemiş	1	1	1.590.000	Akdeniz	İçel	Gerçek	Erkek	Evlü	33	1
232204	Hazır Giyim	1	1	3.550.000	Doğu Anadolu	Erzurum	Gerçek	Kadın	Bekar	28	1
232204	Süt	1	2	575.000	Doğu Anadolu	Erzurum	Gerçek	Kadın	Bekar	28	1
232204	Süt	1	2	575.000	Doğu Anadolu	Erzurum	Gerçek	Kadın	Bekar	28	1
452508	Tekel	1	2	13.750.000	Ege	İzmir	Gerçek	Erkek	Evlü	62	1
452508	Meşrubat ve Kuruyemiş	1	1	1.200.000	Ege	İzmir	Gerçek	Erkek	Evlü	62	1
452508	Meşrubat ve Kuruyemiş	1	1	1.990.000	Ege	İzmir	Gerçek	Erkek	Evlü	62	1
452508	Süt	1	1	890.000	Ege	İzmir	Gerçek	Erkek	Evlü	62	1
452508	Kırtasiye	1	1	5.250.000	Ege	İzmir	Gerçek	Erkek	Evlü	62	1

Tablo 2- Birliktelik Kuralları Analizi İçin Verinin Kodlanması

Müşteri No	Çay-Şeker	Şekerli Ürünler	Bahçe Malzemeleri ve Bitkiler	Baharat Grubu	Bakliyat	Balklar
128402	0	1	0	0	0	0
128402	0	1	0	0	0	0
232204	0	0	0	0	0	0
232204	0	0	0	1	0	1
452508	0	0	0	1	0	0
452508	0	0	0	1	0	0
452508	0	0	0	1	0	1

yanıtlarla SPSS Clementine’i birinci sıraya taşımışlardır. (www.ntvmsnbc.com) Yazılımın üç yıldır üst üste bir numara olması, SPSS’in iş kullamcılarına hitap eden, kullanımı kolay, farklı sistem ve veri yapılarına kolayca entegre edilebilir özelliklerinden kaynaklanmaktadır. Bu özelliklerinden dolayı SPSS Clementine Programı bu çalışmanın uygulama bölümünün gerçekleştirilmesi için seçilmiştir. Tablo 1’de firmaya ait alışveriş verileri görülmektedir.

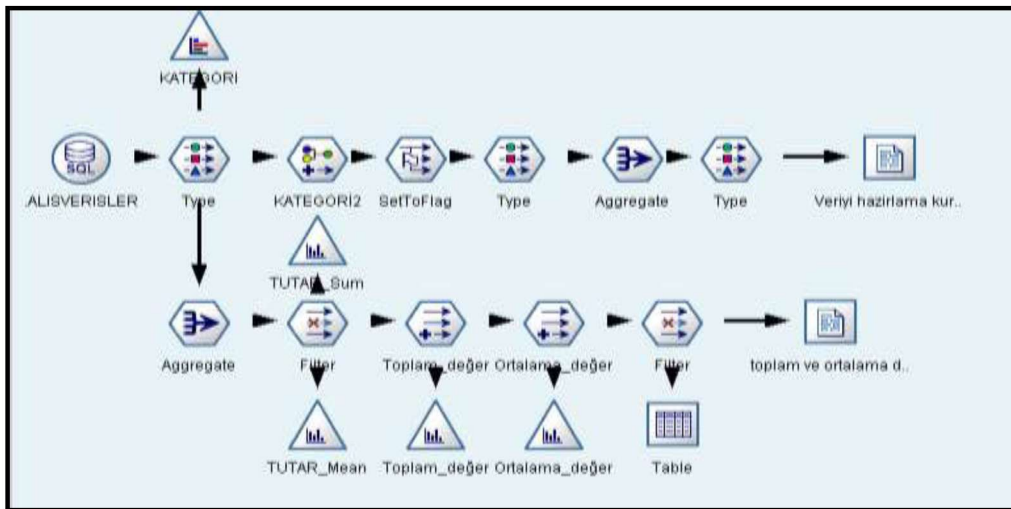
Birliktelik kuralları analizinin yapılabilmesi için verinin kategori bazında 0-1 olarak kodlanması gerekmektedir. Müşterilerin satın aldıkları ürünler için “1”, satın almadıkları ürünler için “0” kullanılmıştır. Bu

uygulamanın gerçekleştirilebilmesi için veri matrisinin devriği alınarak bütün kategoriler ikili alanlara dönüştürülmüştür.(Tablo 2) Bu ikili alanlar müşterinin ziyaretinde o ürünü satın alıp almadığını göstermektedir.

Analize konu değişkenlerin kural çıkarımı için hazırlanması gerekmektedir. Bunun için SPSS Clementine programında gerekli düzenlemeler yapılmış ve bu düzenlemelere ait ara yüz Şekil 1’de verilmiştir.

Veride yer alan ürün kategorileri, ürünlerin satın alınma oranları ve satın alınma frekansları Tablo 3’te görülmektedir.

Şekil 1- Birliktelik Kuralları Analizi için Veriyi Hazırlama

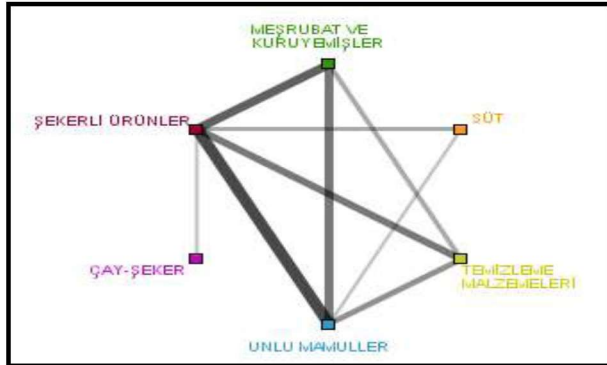


Tablo 3- Ürün Kategorileri

En Çok Satın Alınan Ürünler	Yüzde (%)
Şekerli Ürünler	16,46
Unlu Mamüller	12,25
Meşrubat ve Kuruyemişler	8,45
Temizlik Malzemeleri	6,72
Süt	5,44
Çay-Şeker	4,29
Deterjan	3,76
Peynirler	3,73
Tekel	3,61
Tereyağı-Yoğurt	3,53

Tablo 3 incelendiğinde en çok %16,46 oran ile şekerli ürünlerin satın alındığı görülmektedir. Daha sonra %12,25 oranı ile unlu mamuller ve %8,45 oranı ile meşrubat ve kuruyemişler satın alınmaktadır.

Birliktelik Kurallarının belirlenebilmesi için ilk olarak Web Grafığı ile kategoriler arası ilişkiler incelenmiştir. Kategoriler arasında sıklığı yüksek olan bağıntılar Şekil 2’de verilmiştir. En sık Meşrubat ve Kuruyemişler, Şekerli Ürünler ve Unlu Mamuller satın alınmaktadır.



Şekil-2 Sıklığı Yüksek Olan Bağntılar

Tablo 4- Birliktelik Kuralları Analizi Sonucu

		Destek Oranı (%)	Güven Oranı (%)
Süt			
Temizlik Malzemeleri	Şekerli Ürünler	42,994	96,667
Temizlik Malzemeleri			
Meşrubat ve Kuruyemişler	Şekerli Ürünler	48,567	94,754
Temizlik Malzemeleri			
Unlu Mamüller	Şekerli Ürünler	52,866	93,675
Meşrubat ve Kuruyemişler			
Unlu Mamüller	Şekerli Ürünler	54,936	92,464
Süt	Şekerli Ürünler	54,299	91,496
Temizlik Malzemeleri			
Şekerli Ürünler	Unlu Mamüller	54,777	90,407

Tablo 4’te görülen alanlar şöyle yorumlanmaktadır. Yorumlar ilk üç satırdaki veriler dikkate alınarak yapılmıştır.

Yapılan analizde süt ve temizlik malzemeleri ile şekerli ürünlerin, toplam işlemlerin %42,994’ünde birlikte satın alındıkları görülmüştür. Süt ve temizlik malzemeleri alanların %96,667’si, şekerli ürün de satın almaktadır.

Temizlik malzemeleri ve meşrubat ve kuruyemiş ile şekerli ürünlerin, toplam işlemlerin %48,567’sinde birlikte satın alındıkları görülmüştür. Temizlik malzemesi ile meşrubat ve kuruyemiş alanların %94,754’ü şekerli ürün de almaktadır.

Temizlik malzemeleri ve unlu mamuller ile şekerli ürünlerin, toplam işlemlerin %52,866’sında birlikte satın alındıkları görülmüştür. Temizlik malzemeleri ve unlu mamul alanların %93,675’i, şekerli ürün de almaktadır.

Tablo 4’te güven oranı %90’dan yüksek olan ürünler görülmektedir.

5.SINIFLANDIRMA VE REGRESYON AĞAÇLARI İLE KURAL ÇIKARIMI

Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Trees, C&RT) algoritması veriyi iki alt kümeye ayırmaktadır. Böylece bir sonraki adımda oluşacak olan alt küme, bir öncekinden daha homojen olmaktadır. Bu süreç sonuç bulunana kadar devam eden, kendini tekrarlayan bir süreçtir. C&RT sınıflandırma ve regresyon analizi için kullanılan bir algoritmadır.

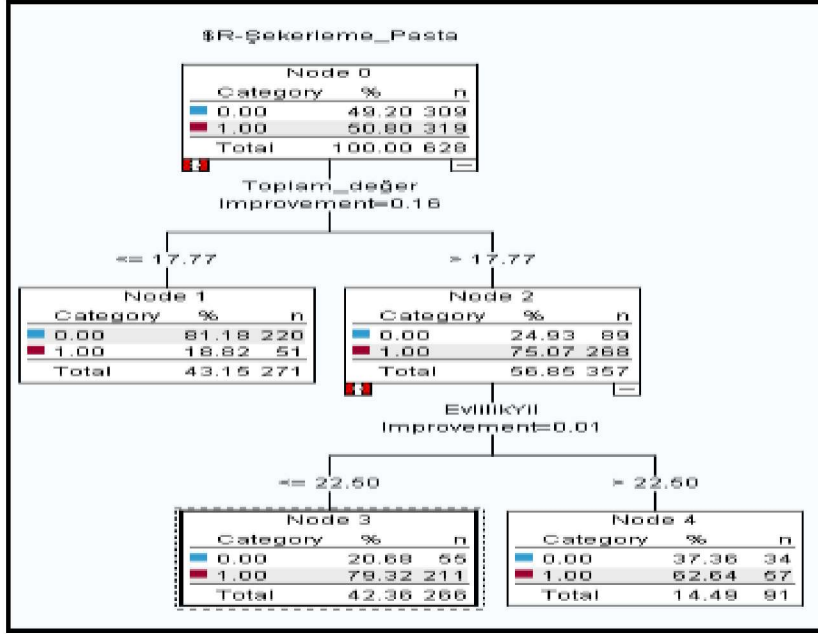
“Şekerli ürünler”, “unlu mamuller”, “meşrubat ve kuruyemiş” ürünlerinin her üçünü birden alan müşteriler için Şekerleme Pasta adında yeni bir alan yaratılmıştır. Bu alanda üç ürünü birden alan müşteriler 1, diğer müşteriler ise 0 ile temsil edilmektedir.

“Şekerli ürünler”, “unlu mamuller”, “meşrubat ve kuruyemiş” ürünlerinin her üçünü birden alan müşterilerin profillerini belirlemek için karar ağacı kullanılmıştır.

Hedef değer : Şekerleme_Pasta olarak tanımlanmıştır.

Tahminleyici alanlar, hedef alan üzerindeki etkinliklerine göre belirlenmiştir. Hedef alan üzerinde en fazla etkili olan iki değişken, toplam alışveriş tutarını gösteren “Toplam değer” ve “Evlilik yılı” değişkenleridir.

“Toplam değer” ve “Evlilik yılı” değişkenleri ile karar ağacı analizi yapılmış ve elde edilen sonuçlar Şekil 3’te sunulmuştur. Diğer değişkenler de modele eklenip karar ağacı analizi yapılmış ve anlamlı sonuçlar elde edilmediği için bu sonuçlar çalışmaya eklenmemiştir.



Şekil 3- C&RT ile Karar Ağacı Analizi Sonuçları
SONUÇ

Şekil 3'te elde edilen sonuçlara göre, müşterilerin %50.80'i Şekerleme_Pasta satın almaktadır. Bu oran alışveriş yapan 628 müşterinin 319'una karşılık gelmektedir. Şekerleme_Pasta satın almayan müşteriler %49.20 oranındadır ve geriye kalan 309 kişiyi oluşturmaktadırlar. Şekerleme_Pasta satın alan kişiler için, toplam satış tutarını gösteren "Toplam değer" değişkeni karar ağacında anlamlı fark yaratan bir değişkendir. Burada hesaplanan 17.77 değeri ayırıcı değerdir.

Yaptıkları alışveriş tutarı 17.77 rakamının altında olan müşterilerin %81.18'i Şekerleme_Pasta satın almamaktadır ve bu oran 220 kişiye karşılık gelmektedir. Yaptıkları alışveriş tutarı 17.77 rakamından büyük olan müşterilerin %75.07'si Şekerleme_Pasta satın almaktadır ve bu oran 268 kişiye karşılık gelmektedir. Karar ağacının bu dalında "Evlilik yılı" önemli bir ayırıcı kriterdir. Evlilik yılının 22.5 değerinden küçük olması ve bu değerden büyük olması şeklinde iki dal oluşmuştur. Evlilik yılı 22.5'ten küçük olan kişilerin %79.32'si Şekerleme_Pasta satın almaktadır ve bu oran 211 kişiye karşılık gelmektedir. Evlilik yılı 22.5'tan büyük olanların %62.64'ü Şekerleme_Pasta satın almaktadır. Oranlar birbirine yakın görünse de, karar ağacında evlilik yılı ayırıcı bir özellik olarak görülmektedir.

Modelin doğruluğu analiz edildiğinde %77.71 oranında doğruluk payı olduğu ve %22.29'luk yanlışlık payı olduğu görülmektedir.

Bu çalışmada Türkiye'de perakende sektöründe faaliyet gösteren büyük bir market zincirinin müşteri verileri ve bu müşterilere ait alışveriş verileri, Birliktelik kuralları kullanılarak analiz edilmiştir. Birliktelik kuralları ve sepet analizi ile, müşterilerin hangi ürünleri hangi ürünlerle birlikte satın aldıkları belirlenmiştir. Daha sonra müşterilerin satın alma davranışını en çok etkileyen değişkenler karar ağaçlarıyla tespit edilmiştir. Analizler sonucunda elde edilen değerler incelendiğinde belirli bir X ürününü alan müşterilerin, belirli bir oranda Y ürününü de satın aldıkları, fakat bunun tersi bir durum olan Y ürününü satın alanların X ürününü satın alanlar ile aynı oranda olmadığı görülmüştür. Bu ve benzeri bilgiler, kampanya düzenlemeleri veya raf dizilişlerinde ve ilişkili ürünlerle ilişkili olmayan ürünlerin promosyon ve satışlarında da kullanılabilir verilerdir.

Çalışma sonucunda elde edilen bilgiler ile, müşteri beklentileri analiz edilerek, satın alma davranışlarına göre farklı ürünleri satın almaları sağlanarak, firmanın kar marjı yükseltilebilecektir. Mevcut müşterilerin yanı sıra hangi müşteri gruplarının hedeflenmesi gerektiği de yine analizler sonucunda elde edilecek faydalı diğer bilgilerdendir.

Sepet analizi ile elde edilen sonuçlar, reklam stratejileri belirlemede, satış işlemini planlamada, katalog dizaynlarında kullanılabilir. Sepet analizi yöntemi, farklı raf dizimlerinin yapılmasına olanak tanır. Birlikte sık olarak alınan ürünler raflarda yakın yerlere dizilebilirler.

KAYNAKÇA

- AKPINAR Haldun, Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, İ.Ü. İşletme Fakültesi Yayını, 1998, s.6.
- ALPAYDIN Ethem, “Zeki Veri Madenciliği”, Bilişim 2000 Eğitim Semineri, 2000, s.9.
- ANGIULLI F., IANNI G. ve PALOPOLI L., “On The Complexity of Inducing Categorical and Quantitative Association Rules”, *Theoretical Computer Science*, Volume 314, Issues1-2, 2004, s.217-249.
- CHEN Y., TANG K., SHEN R. VE HU Y., “Market Basket Analysis in a Multiple Store Environment”, *Decision Support Systems*, Volume 40, Issue 2, 2005, s. 339-354.
- CHEN M., “Ranking Discovered Rules from Data Mining with Multiple Criteria by Data Envelopment Analysis”, *Expert Systems with Applications*, Vol.33, Issue 4, 2007, s. 1110-1116.
- DOLGUN M.Ö. ve ZOR İ., “Bir Alışveriş Merkezinden Yapılan Satışlar İçin Sepet Analizi”, www.spss.com.tr/bilism06_dosyalar/muhsin_ozgur_dolgun.pdf, 2006.
- FAYYAD U., PIATETSKY-SHAPIO G. ve SMYTH P., “Knowledge Discovery and Data Mining: Towards a Unifying Framework”, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-1996)*, Portland, Oregon, 1996.
- FREITAS A., *Data Mining and Knowledge Discovery with Evolutionary Algorithms*, Almanya, Springer Publications, 2002, s.1-2.
- GIUDICI P. ve PASSERONE G., “Data Mining of Association Structures to Model Consumer Behaviour”, *Computational Statistics & Data Analysis*, Vol. 38, Issue 4, 2002, s. 533-541.
- HRUSCHKA H., LUKANOWICZ M. ve BUCHTA C., “Cross-Category Sales Promotion Effects”, *Journal of Retailing and Consumer Services*, Volume 6, Issue 2, 1999, s.99-105.
- LIU L., BHATTACHARYYA S., SCLOVE S., CHEN R. ve LATTYAK W., “Data Mining on Time Series: An Illustration Using Fast-Food Restaurant Franchise Data”, *Computational Statistics & Data Analysis*, Volume 37, Issue 4, 2001, s. 455-476.
- MILD A. ve REUTTERER T., “An Improved Collaborative Filtering Approach for Predicting Cross-Category Purchases Based on Binary Market Basket Data”, *Journal of Retailing and Consumer Services*, Volume 10, Issue 3, 2003, s.123-133.
- MIN H., MIN H. ve EMAM A., “A Data Mining Approach to Developing the Profiles of Hotel Customers”, *International Journal of Contemporary Hospitality*, Vol.14, Issue 6, 2002, s.274.
- MOSS Larissa T., *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision- Support Applications*, Addison Wesley, 2003, s.307-310.
- ÖZEL T., “Finans Sektöründe İstatistik ve Veri Madenciliği Uygulamaları”, SPSS Sunumu, 2003.
- ÖZMEN Ş., “İş Hayatı Veri Madenciliği ile İstatistik Uygulamalarını Yeniden Keşfediyor”, V. Ulusal Ekonometri ve İstatistik Sempozyumu, Adana, 2001.
- RYGIELSKI C., WANG J. ve YEN D.C., “Data Mining Techniques for Customer Relationship Management”, *Technology in Society*, Volume 24, Issue 4, 2002, s. 483-502 .
- SEVER H. ve OĞUZ B., “Veritabanlarında Bilgi Keşfine Formal bir Yaklaşım, Kısım 1: Eşleştirme Sorguları ve Algoritmalar”, *Bilgi Dünyası*, 3(2), 2002.
- SPSS Uygulama Alanları, <http://www.spss.com.tr/uygulamaalanlari.htm>, 22.07.2005.
- TUĞ E. ve BULUN M., “Tıbbi Veri Tabanlarında Gizli Bilgilerin Keşfedilmesi”, www.tbd.org.tr, 03.03.2006.
- VINDEVOGEL B., VAN DEN POEL D. ve WETS G., “Direct and Indirect Effects of Retail Promotions on Sales and Profits in the Do-It-Yourself Market”, *Expert Systems with Applications*, Volume 28, Issue 3, 2005, s. 583-590.
- WANG H. ve WEIGEND A., “Data Mining for Financial Decision Making”, *Journal of Decision Support Systems*, Vol.37, No:4, 2003, s. 148.
- WEI C. ve CHIU I., “Turning Telecommunications Call Details to Churn Prediction: a Data Mining Approach”, *Expert Systems with Applications*, Volume 23, Issue 2, 2002, s. 103-112 .
- YEN S. ve LEE Y., “An Efficient Data Mining Approach for Discovering Interesting Knowledge from Customer Transactions”, *Expert Systems with Applications*, Volume 30, Issue 4, 2006, ss.650-657.
- Yöneylem Araştırması ve Endüstri Mühendisliği Doktora Öğrencileri Kolokyumu, “Hizmet-Kar Zinciri Perspektifinden Müşteri Tabanı Analizi”, Kültür Üniversitesi, (2003).
- “Veri Madenciliğinin Yaygın Uygulama Alanları”, www.spss.com.tr, 11.05.2005.
- “Veri Madenciliğinde Tercih Clementine”, www.ntvmsnbc.com, 16.08.2006.