

MÜŞTERİ PROFİLİ VE ALIŞVERİŞ DAVRANIŞLARINI BELİRLEMEDE KÜMELEME VE BİRLİKTELİK KURALLARI ANALİZİ: PERAKENDE SEKTÖRÜNDE BİR UYGULAMA

Mehpare TİMOR

Istanbul Üniversitesi, İşletme Fakültesi,
Sayısal Yöntemler Anabilim Dalı

Ayşegül EZERÇE

U.Tuğba GÜR SOY

Istanbul Üniversitesi, İşletme Fakültesi,
Sayısal Yöntemler Anabilim Dalı

ÖZET

Bilginin öneminin yıllar boyunca artması ile verinin içindeki anlamlı ve kullanışlı bilginin çıkarılmasını sağlayan Veri Madenciliği tekniği, özellikle işletmelerin başarısı için önemli bir araç haline gelmiştir. Bu çalışmada hazır giyim perakende sektöründe faaliyet gösteren bir firmanın alışveriş kayıtları ile alışverişini gerçekleştiren müşteri verileri ele alınmıştır. Analizde öncelikle “Birliktelik Kuralları Analizi” ile müşterilerin alışveriş alışkanlıkları belirlenmeye çalışılmış, daha sonra “Kümeleme Analizi” ile müşteriler, demografik özellikleri dikkate alınarak bölümlendirilmiştir.

Anahtar Kelimeler: Birliktelik Kuralları, Kümeleme Analizi, Veri Madenciliği, Perakendecilik.

ABSTRACT

Data Mining is a technique, which enables discovering meaningful and useful information in data has been critical factor for the success by the increase of the importance of the knowledge over the years. In this study, shopping data of a confection retail firm with the data of the owners of the shopping is handled and first of all the shopping behavior of customers are determined by using Association Rules which is one of the technique of Data Mining. After that an application is presented by dividing customers into segmentations considering their demographic specifications by using Clustering techniques.

Keywords: Association Rules, Clustering, Data Mining, Retailing.

1. GİRİŞ

Bilgisayar ve iletişim teknolojilerindeki gelişmeler, geçmişe göre çok daha fazla verinin; daha hızlı toplanması, depolanması, işlenmesi ve bilgiye dönüştürülüp yeniden istenen noktalara iletilmesini sağlamıştır. Günümüz iş hayatında, bilgi hacimindeki bu olağanüstü

artış nedeniyle, organizasyonlar elde toplanan bu verilerden nasıl faydalanılabileceği problemi ile karşı karşıya kalmaktadır. Geleneksel sorgu (Query) ve raporlama araçlarının yetersiz kalması nedeniyle, Veri Tabanlarında Bilgi Keşfi-VTBK (Knowledge Discovery in Databases) adı altındaki yeni arayışlara sebep olmuştur. Bu amaçla, VTBK süreci içerisinde, modelin kurulması ve değerlendirilmesi aşamalarından meydana gelen Veri Madenciliği (Data Mining) kullanılmaktadır (Akpınar, 2000, s: 1). Veri madenciliği; veri ambarlarındaki tutulan çok çeşitli verilere dayanarak daha önce keşfedilmemiş bilgileri ortaya çıkarmak, bunları, karar vermek ve eylem planını gerçekleştirmek için kullanma sürecidir (Özmen, 2001, s: 1-6).

Veri madenciliği, eldeki verilerden; üstü kapalı, çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılmasını sağlayan bir tekniktir. Veri madenciliği istatistik, makine bilgisi ve veri tabanları sistemleri ile yakından ilişkilidir.

Veri madenciliğinde; sınıflandırma, kümeleme ve birliktelik kuralları analizi gibi konularda kullanılmak üzere birçok algoritma geliştirilmiştir. Bu yöntemlerin bir çoğu istatistik tabanlıdır. Bu algoritmaların bazıları sadece sınıflandırma veya kümeleme gibi konuları ilgilendirirken, bazıları ise değişik varyasyonlarla birden fazla konuda kullanılabilir. Örneğin, genetik algoritmalar, yapay sinir ağları gerek sınıflandırma ve gerekse kümeleme modellerinde kullanılabilirler; Apriori algoritması ise sadece birliktelik kurallarının belirlenmesinde kullanılan bir algoritmadır (Silahtaroglu 2008, s: 30).

Veri madenciliği teknikleri “Tahmin Edici” ve “Tanımlayıcı Teknikler” olmak üzere ikiye ayrılmaktadır. Tahmin edici tekniklerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. Örneğin, bir banka önceki dönemlerde vermiş olduğu kredilere ilişkin gerekli tüm verilere sahip olabilir. Bu verilerde, bağımsız değişkenler; kredi alan müşterinin özellikleri, bağımlı değişken ise; kredinin geri ödenip ödenmediğidir. Bu verilere uygun olarak kurulan model, daha sonraki kredi taleplerinde müşteri özelliklerine göre verilecek olan kredinin geri ödenip ödenmeyeceğinin tahmininde kullanılmaktadır (Özekeş, 2003, s: 65-82).

Tanımlayıcı tekniklerde ise karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır (Akpınar, 2000, s: 4).

Bu çalışmada, hazır giyim perakende sektöründe faaliyet gösteren bir firmanın alışveriş kayıtları ile alışverişini gerçekleştiren müşteri verileri ele alınarak, Birliktelik Kuralları Analizi aracılığıyla müşterilerin alışveriş alışkanlıkları belirlenmeye çalışılmıştır. Kümeleme Analizi ile müşteriler demografik özelliklerine göre ayrılmıştır.

2. LİTERATÜR ARAŞTIRMASI

Veri madenciliği teknikleri günümüz iş hayatında bir çok alanda başarı ile kullanılmaktadır. Bu uygulama alanlarının başlıcaları; **Pazarlama alanında** müşterilerin satın alma örüntülerinin belirlenmesi, müşterilerin demografik özellikleri arasındaki

bağlantıların bulunması, posta kampanyalarında cevap verme oranının artırılması, mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması, Pazar sepeti analizi, Müşteri ilişkileri yönetimi, müşteri değerlendirme ve satış tahmini; **Bankacılıkta** farklı finansal göstergeler arasında gizli korelasyonların bulunması, kredi kartı dolandırıcılıklarının tespiti, kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi ve kredi taleplerinin değerlendirilmesi; **Sigortacılık alanında** ise yeni poliçe talep edecek müşterilerin tahmin edilmesi, sigorta dolandırıcılıklarının tespiti ve riskli müşteri örüntülerinin belirlenmesi olarak sıralanabilir (SAS, 2009).

Veri madenciliği modelleri,

- Sınıflama (Classification) ve Regresyon (Regression),
- Kümeleme (Clustering),
- Birliktelik Kuralları (Association Rules) ve Ardışık Zamanlı Örüntüler (Sequential Patterns),

olmak üzere üç ana başlık altında incelenmektedir (Akpınar, 2000, s. 4) .

Müşterilerin satın alma davranışlarını tanımlamada kullanılan birliktelik kuralları ve ardışık zamanlı örüntüler, veri madenciliğinde, **Pazar Sepeti Analizi (Market Basket Analysis)** adı altında yaygın olarak kullanılmaktadır. Alışverişlerde müşterinin hangi mal veya hizmetleri satın almaya eğilimli olduğunun belirlenmesi amacıyla kullanılan bu teknik, bir müşteriye daha fazla ürünün satılması yollarından biridir. Birliktelik kuralları müşteri alışveriş davranışlarını belirlemede etkili bir yöntemdir. Brijs ve diğerleri çalışmalarında, ürün yerleştirme kararını geliştirmek için bir yaklaşım önermişlerdir. İlgili çalışmada çapraz satışları arttırmak için sıklıkla alınan ürünler tespit edilmiştir (Brijs, 2004, s: 7-23).

Sepet analizi, perakende sektöründe yaygın olarak kullanılan bir tekniktir. Decker ve Monien'in çalışmasında iki farklı yapay sinir ağı yaklaşımı sunulmuş ve ürün satın almalarına uygunlukları anlatılmıştır. Ele alınan bu iki yöntem Alman perakende mağazalar zincirine uygulanmış, modellerin, satın alma birlikteliklerine ne kadar yakın sonuç verdiği kontrol edilmiştir (Decker, 2003, s: 373-386).

Müşteri profili oluşturma çalışmaları pazarlama sektöründe yaygın olarak kullanılmaktadır. Barber ve diğerlerinin çalışmasında, farklı bir alanda, sağlık sektöründe lazerle göz ameliyatı için müşteri profili çıkarma, bu ameliyat için en uygun hastaların belirlenmesi amacıyla lojistik regresyon analizi gerçekleştirilmiştir (Barber, 2001, s: 32-36).

İnternetin yaygın olarak kullanılması, firmaların bu durumu rekabet avantajına dönüştürmek istemesine yol açmaktadır. Firmalar rakiplerinden farklılaşmak ve rekabet üstünlüğü elde etmek için müşteri hizmetlerini geliştirmek durumundadır. Farklı sebeplerle on-line servisleri kullanan müşteriler söz konusudur Özer'in çalışmasında on-line müzik endüstrisi ele alınmış ve bulanık küme analizi uygulanmıştır. Analiz sonucunda on-line müzik servislerini kullanan potansiyel kullanıcılar arasında homojen grupların bulunduğu tespit edilmiştir. Elde edilen her grup için farklı stratejiler geliştirilmiştir (Özer, 2001, s: 193).

3. KÜMELEME ANALİZİ

Kümeleme analizi birinci amacı “gözlem birimlerini benzer özelliklerine göre gruplamak” olan çok değişkenli istatistik tekniklerden biridir. Kümeleme analizinde elde edilen bir küme içindeki gözlem birimleri, önceden belirlenmiş bir özellik bakımından birbirine benzemektedir. Dolayısıyla elde edilen kümedeki gözlem birimleri homojendir (Hair, 1998, s: 473). Kümeleme analizinde temel hedef, dağınık bir halde bulunan verileri benzerliklerine göre bir araya getirip sınıflandırarak işlenebilir hale getirmektir (Koç, 27.06.2005, s: 2).

Kümeleme modellerinde amaç; küme üyelerinin birbirine çok benzediği, ancak özellikleri birbirinden çok farklı olan kümelerin elde edilmesi ve veri tabanındaki kayıtların bu farklı kümelere bölünmesidir (Akpınar, 2000, s. 4). Kümeleme analizi, araştırma konusu olan birey ya da nesnelere, aralarındaki benzerlikler itibarıyla bir araya getirir. Böylece bu birey ya da nesnelere, çeşitli özellikleri itibarıyla kümelere ayrılırken, kümeler içi homojenlik ve kümeler arası heterojenlik maksimum düzeyde tutulmuş olur. Bir diğer deyişle, bir kümeyi oluşturan bireyler birbirleri ile benzerlerken, diğer kümelerin bireyleriyle benzeşmeyecektir (Nakip, 2003, s: 418).

Kümeleme analizi sınıflandırmaya benzer bir tekniktir. Aradaki en belirgin fark, kümelerin önceden belirlenmemiş olmasıdır. Kümeleme analizinde çoğunlukla yapay sinir ağları ve istatistiksel metodlardan yararlanılmaktadır. Kümeleme analizinin amacı, bir veya birkaç özellik açısından benzer olan nesnelere veya değişkenleri belirlemektir (Aaker, 1971, s: 299).

En yaygın olarak kullanılan kümeleme algoritması, k-ortalamlar yöntemidir.

K-ortalamlar (k-means) yöntemi 1967 yılında Mac Queen tarafından sunulmuştur. Bu yöntem uzun yıllar boyunca pek çok uygulama alanında yoğun olarak kullanılan bir kümeleme algoritmasıdır. Bu algoritmada, k sayıda küme oluşmakta ve her küme içerisinde bulunan verilerin ağırlıklı ortalamaları sonucu bir değer ortaya çıkmaktadır. Küme içerisinde bu değere en yakın olan nokta değeri “küme merkezi” (centroid) olarak kabul edilmektedir (Berkhin, 2002, s: 15). K-ortalamlar yöntemi öncelikle eldeki verileri, kümelerin ortalamalarına göre kümelere ayırır. k küme sayısı kullanıcı tarafından verilir. Burada bahsedilen ortalama, küme merkezidir. Daha sonra gelen her veri merkez noktaya en yakın olduğu kümeye dahil edilir. Eklendiği küme elemanlarının ağırlıklı ortalamaları tekrar hesaplanarak yeni bir küme merkezi değeri bulunur ve bu yeni değer bundan sonraki işlemlerde bu kümeyi temsil eder.

4. BİRLİKTELİK KURALLARI ANALİZİ

Veri tabanlarındaki bilgi miktarı arttıkça birçok kurum ve kuruluş, sahip oldukları bilgiler arasındaki ilişkileri ortaya çıkarma çabası içerisinde girmiştir. Yığınlar halindeki bilgiler arasındaki ilişkiler, kurumlar için önemli sonuçlar doğurabilecek kararların alınmasında

stratejik bir rol oynamaktadır. İlişki analizi veri tabanındaki bir dizi bilgi veya kaydın diğer kayıtlar ile olan bağlantısını açıklayan işlemler dizisidir.

İlişki analizi satış-pazarlamadan, ürün katalog tasarımlarına kadar birçok alanda kullanılmaktadır. Örneğin, herhangi bir ürün satın alınırken, bu ürünün yanında başka bir ürün ya da ürünlerin satın alınması, bu ürünler arasındaki bağlantıyı ortaya koymaktadır. Bu bağlantıların bir kural olarak belirlenmesi ise birliktelik kuralları analizi konusuna girer. Literatürde bu tür çalışmalara “Pazar Sepeti Analizi” adı da verilmektedir. Pazar sepeti analizi müşterilerin alışveriş alışkanlıklarının veri tabanındaki bilgiler aracılığıyla ortaya çıkarılması işlemidir. Bu işlemlerin sonuçları alışveriş merkezindeki ürünlerin yerleştirilmesi, market alanının tasarımı ve markette birlikte sergilenecek, satılacak ürünlerin belirlenmesine yardımcı olur (Silahtaroglu, 2008, s: 83).

Birliktelik kuralları olaylar arasındaki probabilistik korelasyonu tanımlamaktadır. Olaylar arasındaki korelasyon sık sık beraber gözlenen olaylardır. Herhangi bir veri tabanında birliktelik kurallarının tanımlanması, veri tabanı bilgi keşfi sürecinin ilk adımudur. Birliktelik kurallarıyla ilgili problem, birliktelik kurallarını belirlemede bir eşik değeri bulmaktır. İlginç birliktelik kurallarından ilginç olmayanları ayırt edebilmek için de bazı ölçütlerin belirlenmesi gereklidir. Bu ölçütler destek ve güven oranlarıdır.

Birliktelik kurallarının amacı, kullanıcı tarafından belirlenen minimum destek ve güven oranlarını sağlayan kuralların bulunmasıdır. Anlamlılığı destek ve güven değerleri ile ölçülen birliktelik kuralları, "X nesnesini alan bir müşterinin muhtemelen Y nesnesini de alması" tipindeki kuralların tanımlanmasını amaçlamaktadır.

Destek Oranı: $P(X \text{ ve } Y) = X \text{ ve } Y \text{ Mallarını Satın Almış Müşteri Sayısı} / \text{Toplam Müşteri Sayısı}$

Güven Oranı: $P(X/Y) = P(X \text{ ve } Y) / P(Y) = X \text{ ve } Y \text{ Mallarını Satın Almış Müşteri Sayısı} / Y \text{ Malını Satın Almış Müşteri Sayısı.}$

formülleri ile hesaplanmaktadır.

Destek oranı, veride bu bağıntının ne kadar sık olduğunu; güven oranı 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, 2009, s: 9).

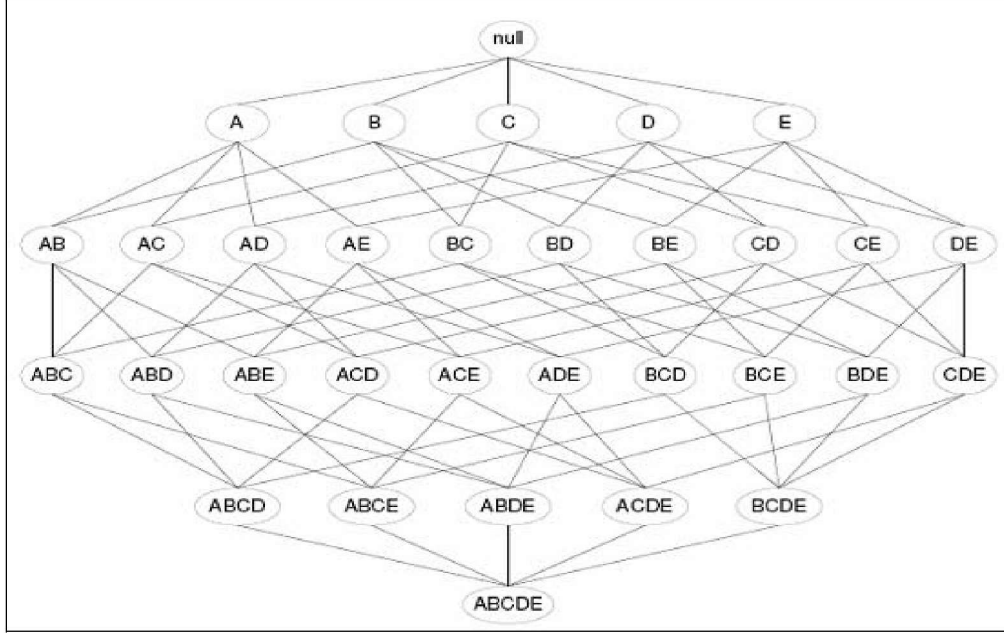
Bu durum çok bilinen bir örnek ile açıklanabilir: “Bira içeren %30 alışverişin, %2 si aynı zamanda çocuk bezi de içermektedir.” Burada %30 güven seviyesini, %2 ise bu güven seviyesine olan desteği belirtmektedir (Pande, 01.01.2007).

Birliktelik kuralı analizinde kullanılan algoritmalarda minimum güven ve destek oranlarını sağlayan birliktelik kuralı çıkarım problemi iki adıma bölünmüştür.

Birinci adım, kullanıcı tarafından belirlenmiş minimum destek kıstasını sağlayan ürün kümelerinin bulunmasıdır. Bu kümelere sık geçen öge kümesi adı verilmektedir. Verilen örnekte N adet ürün (öge) var ise, potansiyel olarak 2^N adet sık geçen öge kümesi

olabilir. Bu adımda üstel arama uzayını etkili biçimde tarayarak sık geçen öge kümelerini bulan etkili yöntemler kullanılmalıdır.

ŞEKİL 1: 2^N Kümesi (Tan, 2005, s: 332)



İkinci adım ise sık geçen öge kümeleri kullanılarak minimum güvenlik kriterini sağlayan birliktelik kurallarının bulunmasıdır. Bu adımdaki işlem şöyle yapılmaktadır: Sık geçen her l öge kümesi için, boş olmayan l 'in tüm alt kümeleri üretilir. l 'nin boş olmayan alt kümeleri a ile gösterilsin. Her a kümesi için $a \rightarrow (1 - a)$ gerektirmesi, l kümesinin destek ölçütünün a kümesinin destek ölçütüne oranı minimum güvenilirlik eşiği ölçütünü sağlıyorsa $a \rightarrow (1 - a)$ birliktelik kuralı olarak üretilir. Minimum destek eşiğine göre üretilen çözüm uzayında, minimum güvenilirlik eşiğine göre taranarak bulunan birliktelikler kullanıcının ilgilendiği ve potansiyel olarak önemli bilgi içeren birlikteliklerdir (Rakesh, 2003, s:2).

Birliktelik sorgusu algoritmalarının performansını belirleyen adım birinci adımdır. Sık geçen öge kümeleri belirlendikten sonra, birliktelik kurallarının bulunması düz bir adımdır (Rakesh, 1994, s: 487-499).

5. UYGULAMA

Perakende sektörü, ulusal ve uluslararası üreticilerin ürünlerini tüketicilerle buluşturduğu büyük bir hizmet platformudur. Perakendecilik, mal ve hizmetlerin doğrudan doğruya son tüketiciye pazarlanmasıyla ilgili faaliyetlerin bütünüdür. Perakende kuruluşları; Gıda Maddeleri, Tütün, Bakkaliye Perakendecileri, Mensucat, Giyim Eşyası ve Deri Eşyası Perakendecileri, Mobilya ve Ev Eşyası Perakendecileri, Madeni Eşya, Cam, Porselen ve

Elektrikli Ev Eşyaları Perakendecileri, İlaç ve İtiryat Perakendecileri, Kereste ve İnşaat Malzemesi Perakendecileri, Makine ve Yedek Parça Perakendecileri, Büyük Mağazalar, Kitap ve Kırtasiye Perakendecileri, Mücevherat, Antika, Oyuncak Perakendecileri şeklinde sıralanmaktadır (Sait, 31.12.2005).

Bu çalışmada, Türkiye'nin en büyük hazır giyim perakende işletmelerinden birinin 2005 yılının Ocak, Şubat, Mart, Nisan ve Mayıs aylarına ait müşteri alışveriş kayıtları kullanılarak perakende mağazacılıkta hazır giyim sektörüne ait analizler yapılmıştır. Uygulama ve analiz sonuçlarına aşağıda yer verilmiştir.

5.1. Analizlerin Uygulanması

Bu uygulamada kullanılan veri setinde gizlilik prensiplerine bağlı kalınarak müşteri, fatura ve mağaza numaraları belirli algoritmalar kullanılarak değiştirilmiştir. Uygulamada kullanılan veri seti belirtilen dönem içinde alışveriş yapmış 5.370 müşterinin 17.164 faturasına ait 42.431 alışveriş kaydından oluşmaktadır. 336'sı sıklıkla alınan 990 farklı mal grubuna sahip olan işletmenin, bu döneme ait alışveriş kayıtlarından sıklıkla kullanılan 281 mal grubundan en çok 123 tanesi tercih edilmiştir. Veri setini oluşturan 5.370 müşterinin 4.624'ü erkek, 746'sı kadındır.

Fayyad ve diğerleri veri tabanlarında bilgi keşfi aşamalarını "Veriyi anlama, hedef veri seti oluşturma, veri temizleme ve önileme, veri indirgeme, amaçları seçme, model analizi ve hipotez seçme, veri madenciliği, örüntüleri yorumlama ve bilgi üzerinde harekete geçme" olarak dokuz aşamada incelemektedir (Fayyad, 1996, s: 37).

Bu uygulamada, veri madenciliği sürecinde, The Cross-Industry Standard Process for Data Mining (CRISP-DM) konsorsiyumu tarafından belirlenmiş standart veri madenciliği süreci izlenmiştir.

5.1.1. Problemin Tanımlanması

Türkiye'nin en büyük hazır giyim perakende işletmelerinden birinden alışveriş yapan müşterilerin alışveriş kayıtları veri seti olarak ele alınmış, bu veri setinden hareketle Birliktelik Kuralları Analizi ile müşterilerin alışveriş bağıntıları tespiti ve Kümeleme Analizi ile bu verilerin ayrıntılı olarak analiz edilmesi amaçlanmıştır. Uygulamada SPSS Clementine programının 11.1 versiyonu kullanılmıştır.

5.1.2. Verinin Hazırlanması

Verinin hazırlanması aşamasında, veri setindeki veriler kullanıma hazır hale getirmek amacı ile birleştirilmiş, temizlenmiş ve gerekli dönüşümler yapılmış olup, aşağıda verilerin hangi aşamalardan geçirilerek düzenlenmiştir.

5.1.2.1 Verinin Elde Edilmesi

İşletmenin amacına yönelik problem tanımlandıktan sonra probleme yanıt verecek çözümü sağlayacak sonuçlar için veri madenciliği sürecinde ilk adım, verilerin elde edilmesi yani veriye ulaşılmasıdır. Bu aşamada Türkiye'nin en büyük hazır giyim perakende işletmelerinden birinin Türkiye'nin çeşitli illerinde yer alan mağazalarının kullanmakta

oldukları Java programlama diliyle yazılmış mağazacılık uygulamasının on-line olarak Oracle veri tabanını beslediği veriler kullanılmıştır.

Veri seti, işletmenin özel kartlarından herhangi birine sahip olan ve olmayan müşterilerinin 2005 yılının ilk beş ayına ait alışveriş kayıtlarından oluşmaktadır. Veri seti alışveriş sırasında anlık olarak tek bir merkezde toplanmış ve sonrasında bir veri ambarına aktarılmış verilerden oluşmaktadır. Veriler Access veri tabanında tek tablo halinde elde edilmiştir ve görüntüsü Tablo 1'de sunulmuştur.

TABLO 1: Verinin Access Veri Tabanı Görüntüsü

MAĞAZA	musid	KAYNAK	C	CINSİYET	MAL_GRUBU	FATURA_HARF	FATURA_NO	TARİH	MIKTAR	TUTAR	İSKONTO	BEDEN_ID	MUSDE	MUSDETA	MUS	MUSD	MUSDETA	E	M
7	546300	30	E	0210	Z		242910	2/4/2005 9:40:35 PM	1	38.98	0.0		ERKEK	1962-10-12	1	EVLU	YEŞİLKÖY	9	9
14	200	25	K	0286	X		519345	4/28/2005 6:18:38 PM	1	75.42	0.000000		ERKEK	1967-06-09	1	EVLU	MUDANYA	9	9
36	297	11	K	0013	K		226226	1/28/2005 6:29:35 PM	1	49.15	0.000000		ERKEK	1966-09-11	1	EVLU	KARŞIYAKA	9	9
36	297	10	E	0120	K		226270	2/6/2005 12:34:37 PM	1	50.85	0.000000		ERKEK	1966-09-11	1	EVLU	KARŞIYAKA	9	9
36	297	11	K	0013	K		226278	2/6/2005 12:34:37 PM	1	33.9	0.000000		ERKEK	1966-09-11	1	EVLU	KARŞIYAKA	9	9
2	1355	21	K	0200	X		121341	2/10/2005 8:20:19 PM	1	19.41	0.000000		ERKEK	1972-09-24	4	EVLU	UMİTKÖY	9	9
3	1355	11	E	0200	X		54329	1/3/2005 12:52:48 PM	1	21.1	0.000000		ERKEK	1972-09-24	4	EVLU	UMİTKÖY	9	9
2	1355	12	K	0013	X		122944	2/24/2005 0:55:50 PM	1	224.50	0.000000		ERKEK	1972-09-24	4	EVLU	UMİTKÖY	9	9
5	1466	10	E	0016	Z		760759	1/18/2005 0:59:25 PM	2	26.1	0.000000		ERKEK	1959-02-13	1	EVLU	BAKIRKÖY	9	9
5	1466	10	E	0016	Z		760759	1/18/2005 8:58:25 PM	1	4.58	0.000000		ERKEK	1959-02-13	1	EVLU	BAKIRKÖY	9	9
5	1466	11	K	0016	Z		760759	1/18/2005 8:58:25 PM	3	29.24	0.000000		ERKEK	1959-02-13	1	EVLU	BAKIRKÖY	9	9
30	1658	21	K	0286	X		125688	2/24/2005 3:19:44 PM	1	30.51	0.000000		ERKEK	1955-11-21	1	EVLU	HADIMKÖY	9	9
30	1658	11	K	0206	X		127675	4/22/2005 2:36:05 PM	1	122.80	0.000000		ERKEK	1955-11-21	1	EVLU	HADIMKÖY	9	9
30	1658	21	K	0241	X		124178	2/12/2005 4:26:16 PM	1	143.22	0.000000		ERKEK	1966-11-21	1	EVLU	HADIMKÖY	9	9
30	2609	10	E	0120	X		122377	1/30/2005 9:30:26 PM	1	83.05	0.000000		ERKEK	1967-02-15	14	EVLU	ÇORLU	9	9
30	2609	10	E	0210	X		122377	1/30/2005 9:30:26 PM	1	95.76	0.000000		ERKEK	1967-02-15	14	EVLU	ÇORLU	9	9
30	2609	10	E	0120	X		122377	1/30/2005 9:30:26 PM	1	74.50	0.000000		ERKEK	1967-02-15	14	EVLU	ÇORLU	9	9
30	2609	10	E	0210	X		122377	1/30/2005 9:30:26 PM	1	83.05	0.000000		ERKEK	1967-02-15	14	EVLU	ÇORLU	9	9
30	2609	10	E	0120	X		122377	1/30/2005 9:30:26 PM	1	95.76	0.000000		ERKEK	1967-02-15	14	EVLU	ÇORLU	9	9
30	2609	10	E	0120	X		122377	1/30/2005 9:30:26 PM	1	56.08	0.000000		ERKEK	1967-02-15	14	EVLU	ÇORLU	9	9
17	3715	30	E	0200	X		595067	2/16/2005 8:42:45 PM	2	11.02	0.000000		ERKEK	1959-11-15	14	EVLU	ARAKLI	9	9
17	3715	30	E	0200	X		595069	2/10/2005 6:02:55 PM	1	5.51	0.000000		ERKEK	1959-11-15	14	EVLU	ARAKLI	9	9
1	3970	10	E	0120	X		18769	3/26/2005 7:07:42 PM	1	41.95	0.000000		ERKEK	1976-09-21	21	EVLU	BEŞİKTAŞ	9	9
1	3978	10	E	0120	Z		164531	3/26/2005 7:07:42 PM	1	50.85	0.000000		ERKEK	1976-09-21	21	EVLU	BEŞİKTAŞ	9	9
50	3983	11	K	0013	X		575963	2/13/2005 4:00:31 PM	1	33.9	0.000000		ERKEK	1971-09-23	25	EVLU	GAYRETTEPE	9	9
38	4058	09	-	0580	X		77219	2/6/2005 12:24:04 PM	1	19.07	0.000000		ERKEK	1967-02-24	12	EVLU	YEŞİLKÖY	9	9
11	5060	09	-	0580	X		626862	4/22/2005 6:49:41 PM	1	55.00	0.000000		ERKEK	1933-01-10	14	EVLU	ŞİŞLİ	9	9
11	5060	10	-	0527	O		34211	2/22/2005 1:44:07 PM	1	37.29	0.000000		ERKEK	1933-01-10	14	EVLU	ŞİŞLİ	9	9
8	5114	24	K	0220	X		41783	4/5/2005 4:52:22 PM	1	21.19	0.000000		ERKEK	1961-06-11	42	EVLU	BEŞİKTAŞ	9	9
1	5130	10	E	0210	Z		163784	3/17/2005 1:14:18 PM	1	25.42	0.000000		ERKEK	1958-12-09	32	EVLU	ŞİŞLİ	9	9
8	5156	10	K	0010	X		42909	5/1/2005 1:22:51 PM	1	46.61	0.000000		ERKEK	1955-06-11	21	EVLU	9	9	
7	5169	30	E	0120	Z		244455	3/11/2005 1:05:55 PM	1	44.92	0.000000		ERKEK	1973-02-06	1	EVLU	BAYRAMPAŞA	9	9

5.1.2.2. Verinin Birleştirilmesi ve Temizlenmesi

Veri madenciliğinde veri kümesinin büyüklüğünden kaynaklanan en fazla zaman alıcı aşama, verilerin ön işlemden geçirilmesi aşamasıdır. Veri madenciliği uygulamalarında kaynakların %80'i verilerin ön işlemden geçirilmesi ve temizlenmesi süreçleri için harcanmaktadır (Piramuthu, 2004, s: 483-494).

Uygulamada kullanılan veri setini oluşturan veriler, söz konusu işletmenin Türkiye'nin çeşitli illerindeki mağazalarının 2005 yılının ilk beş ayına ait alışveriş kayıtlarında bulunan ve analiz için gerekli alanları tam olan verilerden oluşmaktadır. Veri seti ürün (barkod) bazında değil, ürünün bir üst katmanı olup işletme tarafından *Mal Grubu* olarak adlandırılan grup bazında hazırlanmıştır. Her bir analiz için gereksiz sütunlar elenmiş, sadece gerekli sütunlardan oluşan veri setleri tanımlanmıştır.

Veri setinde numarası bulunduğu halde tanımı bulunmayan Mağaza No sütunu ve İskonto Tutarı sütunu analizde yararlı olmayacağı düşünülerek veri setinden çıkarılmıştır. Buna ek olarak, Ürün Beden No ve Müşteri Beden No sütunları tam dolu ve doğru olmadığı

görülerek veri setinden çıkarılmıştır. Cinsiyeti, Doğum Tarihi, Medeni Durumu, Meslek Numarası, İlçe Numarası, İl Numarası bilgileri eksik olan hiçbir müşteri alışveriş kaydı, uygulama veri seti içerisine dahil edilmemiştir.

Veri setindeki sütünlarda verilerin eksik değerlerinin bulunup bulunmadığını tespit etmek için SPSS Clementine Programının “Data Quality” yani Veri Kalitesi aracı kullanılarak verilerin doluluk oranları kolaylıkla gözlenebilmektedir. Tablo 2’de görüldüğü gibi “Müşteri İlçe No”, “Müşteri Beden” ve “Müşteri Meslek Adı” alanlarındaki veriler eksiktir.

TABLO 2: Verinin Kalitesi

Field	Filter	Field
Mağaza ID	→	Mağaza ID
Müşteri ID	→	Müşteri ID
Kaynak ID	→	Kaynak ID
Cinsiyet ID	→	Cinsiyet ID
Mal Grubu ID	→	Mal Grubu ID
Fatura Harf	→	Fatura Harf
Fatura No	→	Fatura No
Tarih	→	Tarih
Miktar	→	Miktar
Tutar	→	Tutar
İskonto Tutarı	→	İskonto Tutarı
Beden ID	→	Beden ID
Müşteri Cinsiyeti	→	Müşteri Cinsiyeti
Müşteri Doğum Tarihi	→	Müşteri Doğum Tarihi
Müşteri Meslek ID	→	Müşteri Meslek ID
Müşteri Meslek Adı	→	Müşteri Meslek Adı
Müşteri Medeni Durumu	→	Müşteri Medeni Durumu
Müşteri İlçe Adı	→	Müşteri İlçe Adı
Müşteri Şehir ID	→	Müşteri Şehir ID
Müşteri Ülke ID	→	Müşteri Ülke ID

Verinin analize hazırlanmadan önceki hali Tablo 3’te görülmektedir. Tabloda; “Mağaza ID, Müşteri ID, Kaynak ID, Cinsiyet ID, Mal Grubu ID, Fatura Harf, Fatura No, Tarih, Miktar, Tutar, İskonto Tutarı, Beden ID, Müşteri Cinsiyeti, Müşteri Meslek Adı, Müşteri Medeni Durumu, Müşteri İlçe Adı, Müşteri Şehir ID, Müşteri Ülke ID” olmak üzere 21 adet değişken bulunmaktadır.

TABLO 3: Verinin Düzenlenmeden Önceki Hali

Table (21 fields, 42,431 records)								
	musid	KAYNAK_ID	CINSIYET_ID	MAL_GRUBU_ID	FATURA_HARF	FATURA_NO	TARİH	MIKTAR
1	200.000	25	K	0286	X	519345	2005-04-28 18:18:38	1.000
2	200.000	21	K	0172	X	519927	2005-05-20 15:27:22	1.000
3	200.000	25	K	0109	X	519997	2005-05-25 12:10:16	1.000
4	200.000	25	K	0286	X	519345	2005-04-28 18:18:38	1.000
5	200.000	21	K	0172	X	519927	2005-05-20 15:27:22	1.000
6	200.000	21	K	0172	X	1177	2005-05-31 10:58:07	1.000
7	200.000	21	K	0172	X	519345	2005-04-28 18:18:38	1.000
8	200.000	21	K	0170	X	519345	2005-04-28 18:18:38	1.000
9	297.000	11	K	0170	K	550389	2005-05-13 19:52:19	1.000
10	297.000	11	K	0171	K	226526	2005-01-29 18:29:35	1.000
11	297.000	11	K	0013	K	226526	2005-01-29 18:29:35	1.000

SPSS Clementine programında veri setinde eksik değerlerin olup olmadığını anlamak için “Data Audit” aracı kullanılmaktadır. Bu sayede doluluk oranı ve geçerli kayıt sayısı kolaylıkla görülmektedir. Tablo 4’te “Müşteri Sempti”, “Meslek” ve “Yaş” alanlarının eksik olduğu görülmektedir. Müşteri sempti alanı %99.553 oranında, müşteri mesleği alanı %99.739 ve yaş alanı %99.981 oranında doludur. Diğer alanlarda ise eksik kayıt olmadığı görülmektedir. Eksik olan kayıtlar için yapay kayıtlar eklenmiştir.

TABLO 4: Veri Denetleme

Data Audit of [8 fields] #1									
Complete fields (%): 62.5% Complete records (%): 99.27%									
Field	Type	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	
musid	Range	0	0 None	Never	Never	fixed	100	5370	
Müşteri_Cins...	Flag	--	--	Never	Never	fixed	100	5370	
Medeni_Dur...	Flag	--	--	Never	Never	fixed	100	5370	
Müşteri_Sempti	Set	--	--	Never	Never	fixed	99.553	5346	
Meslek	Set	--	--	Never	Never	fixed	99.739	5356	
Fatura_id	Set	--	--	Never	Never	fixed	100	5370	
Yaş	<Default>	--	--	Never	Never	fixed	99.981	5369	
Müşteri_Şehri	Set	--	--	Never	Never	fixed	100	5370	

5.1.2.3 Veriyi Dönüştürme

“Doğum Tarihi” alanı verilerin toplandığı yıl olan 2005’ten çıkarılıp müşteri yaşları hesaplanarak “Yaş” alanı oluşturulmuştur. Satış kaydı olan 123 mal grubu birbiri ile ilişkili mal grubu ve/veya mal grupları ile birleştirilerek mal grubu sayısı 41’e indirgenmiştir. Böylelikle daha anlamlı sonuçlar elde edilmiştir. Birleştirilerek gruplanan mal gruplarının görüntüsü Tablo 5’te verilmiştir.

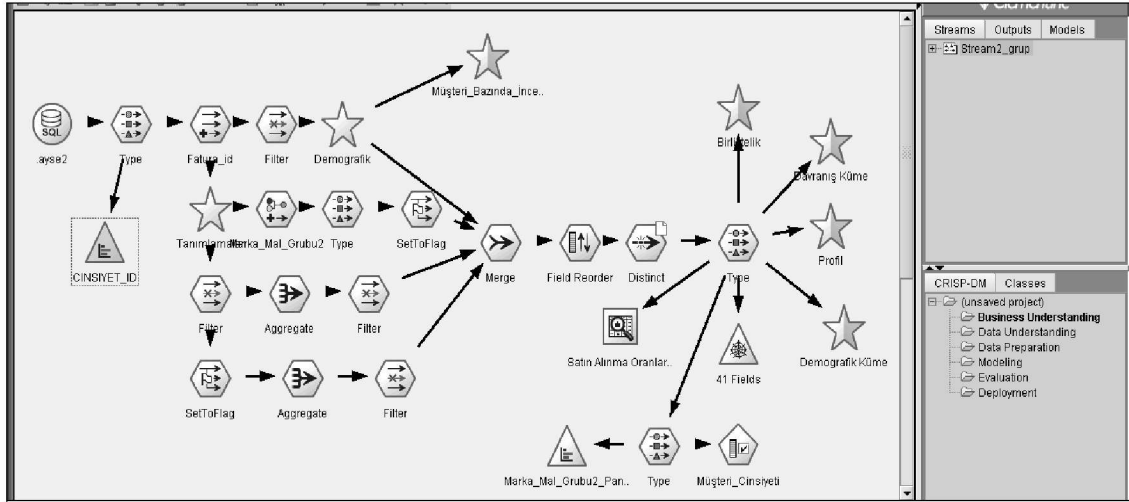
TABLO 5: Gruplanmış Mal Grupları

Field	Proportion ...	%	Count
Marka_Mal_Grubu2_Ayakkabı		22.71	495
Marka_Mal_Grubu2_Bere_Şapka		3.26	71
Marka_Mal_Grubu2_BLUZTUNİK		19.22	419
Marka_Mal_Grubu2_Bornoz_Havlu		2.2	48
Marka_Mal_Grubu2_Çanta_Cüzdán		17.48	381
Marka_Mal_Grubu2_CEKET		16.7	364
Marka_Mal_Grubu2_ÇORAP		11.19	244
Marka_Mal_Grubu2_Deri_ürünleri		8.58	187
Marka_Mal_Grubu2_DÖPIYES		0.41	9
Marka_Mal_Grubu2_ELBİSE		7.61	166
Marka_Mal_Grubu2_ERKEK TAKIM		7.89	172
Marka_Mal_Grubu2_Eşofman		2.8	61
Marka_Mal_Grubu2_ETEK/P.ETEK		15.83	345
Marka_Mal_Grubu2_Ev_Tekstili_Aksesuar		4.04	88
Marka_Mal_Grubu2_GELİNLİK		0.05	1
Marka_Mal_Grubu2_GÖMLEK		27.2	593
Marka_Mal_Grubu2_Hediyeilik		1.1	24
Marka_Mal_Grubu2_İç_Çamaşırı		2.71	59
Marka_Mal_Grubu2_Jean		9.91	216
Marka_Mal_Grubu2_KAŞKOL		3.76	82
Marka_Mal_Grubu2_KEMER/KEMER AKSESU...		13.94	304
Marka_Mal_Grubu2_KIYAFET/DİĞER		0.05	1
Marka_Mal_Grubu2_KOL DÜĞMESİ		3.07	67
Marka_Mal_Grubu2_Kozmetik		1.51	33
Marka_Mal_Grubu2_Kravat_Papyon		16.79	366
Marka_Mal_Grubu2_KUMAŞ		0.92	20
Marka_Mal_Grubu2_KuruÇiçek_Potpori		0.41	9
Marka_Mal_Grubu2_Mayo_Bikini_Pareo		3.07	67
Marka_Mal_Grubu2_MENDİL/POŞET		5.87	128
Marka_Mal_Grubu2_Mont		13.12	286
Marka_Mal_Grubu2_Pantolon_Şort		29.82	650
Marka_Mal_Grubu2_PJAMA/GECELİK		0.55	12
Marka_Mal_Grubu2_Şal_Eşarp_Fular		15.18	331
Marka_Mal_Grubu2_ŞEKERLEME_Çikolata		4.45	97
Marka_Mal_Grubu2_SMOKİN TAKIM		0.46	10
Marka_Mal_Grubu2_SWEATSHIRT		10.78	235
Marka_Mal_Grubu2_Takı ve Aksesuar		16.38	357
Marka_Mal_Grubu2_TAYYÖR		1.51	33
Marka_Mal_Grubu2_TRİKO		37.84	825
Marka_Mal_Grubu2_T-SHIRT/BODY/BUSTİER		36.93	805
Marka_Mal_Grubu2_YELEK		0.6	13

Tablo 5'te gruplanmış ürünlerin satın alınma oranları ve satın alınma frekansları görülmektedir. %37.84 ile en çok alınan ürün grubunun Triko olduğu görülmektedir. Triko ürün grubunu %36.93 ile T-Shirt/Body/Büstiye gruplanmış ürün grubu izlemektedir. Daha sonrasında %29.82 oranıyla Pantolon/Şort gruplanmış ürün grubu gelmektedir.

Analizde kullanılacak 21 değişken İstatistik Analiz, Birliktelik Kuralları ve Kümeleme analizi için hazırlanması gerektiğinden, SPSS Clementine programında gerekli düzenlemeler yapılarak model Tablo 6'da görüldüğü gibi kurulmuştur.

Tablo 6'da görülen Birliktelik Kuralları Analizi modeli kurulduğunda, SPSS Clementine programı fatura bazında alışveriş tablosu oluşturmaktadır. Oluşturulan bu tabloda tüm mal grupları faturalara eklenmektedir. Faturada yer alan mal grupları "T" True, alınmayan mal grupları "F" False olarak kodlanmaktadır. Bu alanlar müşterilerin alışverişlerinde o ürünü satın alıp almadığını göstermekte ve aynı zamanda birlikte alınan ürünlerin belirlenmesi için veriyi hazırlamaktadır.

TABLO 6: Birliktelik Tespiti, Kümeleme Analizi ve Kural Çıkarımı için Modellerin Kurulması

Veride yer alan değişkenlerin istatistiksel değerleri için SPSS Clementine programının “Data Audit” aracı kullanılmış olup istatistiksel sonuçlar Tablo 7’de verilmiştir. Tabloda değişkenlere ait grafikler, değişkenlerin tipleri, değişkenlerin minimum ve maksimum değerleri, ortalamaları, standart sapmaları, çarpıklık değerleri, değişkenlerin aldıkları toplam değer sayıları ve geçerli kayıt sayıları görülmektedir.

Tabloda veri tipi alanına bakıldığında SPSS Clementine tarafından değişkenler için “Range”, “Set” ve “Flag” olmak üzere üç tanımlama yapıldığı görülmektedir. “Range” değişkenin sürekli olduğunu, “Set” değişkenin kategorik olduğunu, “Flag” ise değişkenin iki şıklı değere (Örneğin cinsiyet alanı için kadın/erkek şıklarının bulunması) sahip olduğunu göstermektedir.

TABLO 7: Müşteri Demografik Özellikleri İçin Hazırlanan Verinin Değişkenlerinin Dağılımı ve Temel İstatistikleri

Field	Sample Graph	Type	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
musid		Range	200.000	731317.000	204226.809	194693.955	1.443	--	5370
Müşteri_Cinsiyeti		Flag	--	--	--	--	--	2	5370
Medeni_Durumu		Flag	--	--	--	--	--	2	5370
Müşteri_Semti		Set	--	--	--	--	--	--	5346
Meslek		Set	--	--	--	--	--	64	5356
Fatura_id		Set	--	--	--	--	--	--	5370
Yaş		Range	1	107	43.576	11.021	0.405	--	5369
Müşteri_Şehri		Set	--	--	--	--	--	6	5370

Range olarak tanımlanan Yaş değişkenine bakıldığında, grafiğin normal dağıldığı gözlemlenmektedir. Müşterilerin yaş aralığına bakıldığında 25 ile 65 yaş arasında buldukları, yaş ortalamasının ise 43 olduğu görülmektedir.

Cinsiyet değişkeni incelendiğinde 5.370 müşterinin %86.11'inin erkeklerden, %13.89'unun ise kadınlardan oluştuğu görülmektedir.

Alışveriş kayıtları incelendiğinde yapılan 42.431 alışverişin %55.75'inde kadın ürünleri alındığı ve bu alışverişlerin çoğunlukla erkek müşteriler tarafından yapıldığı görülmüştür.

Medeni Hal değişkeni incelendiğinde müşterilerin %99.89 oranla evli olduğu görülmektedir. Çalışmada kullanılan 42.431 alışveriş kaydını yaratan 5.370 müşterinin 5.364'ünün evlilerden, 6'sının ise bekarlardan oluştuğu görülmektedir.

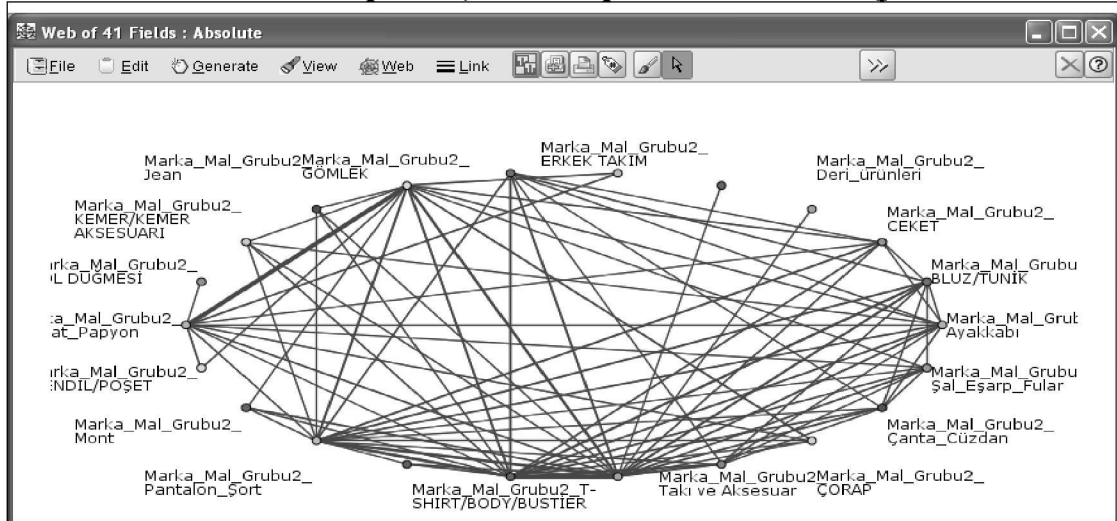
Müşterilerin yaşadıkları şehirler incelendiğinde, müşterilerin çoğunluğunun İstanbul'un Avrupa yakasından olduğu görülmektedir. Meslek alanı incelendiğinde ise müşterilerin çoğunluğunun doktor olduğu görülmektedir.

5.1.3. Modelin Kurulması

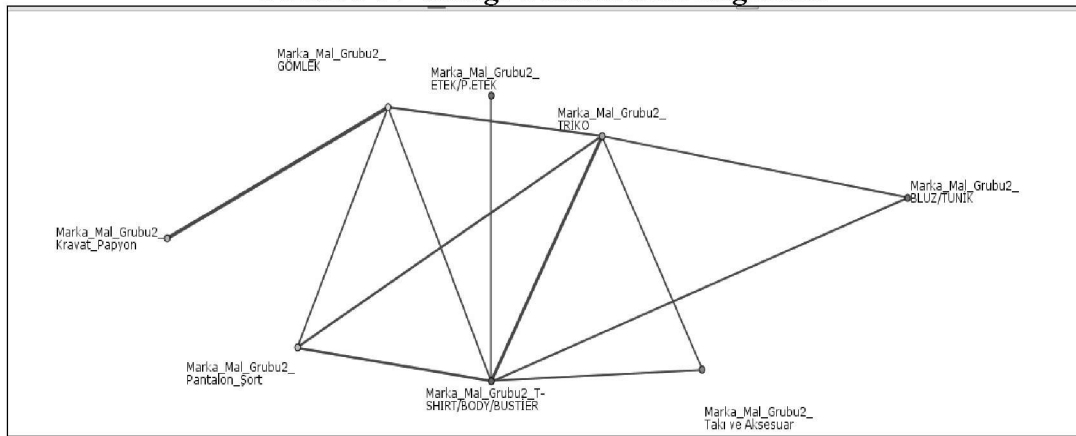
Veri hazırlama ve model kurma aşamaları, en iyi çözümü elde eden modelin bulunmasına kadar tekrarlanan bir süreçtir. Her uygulamada kullanılacak birden çok teknik vardır ve önceden hangisinin en başarılı olacağını kestirmek olası değildir. Bu yüzden öğrenme kümesi üzerinde L değişik teknik kullanılarak L tane model oluşturulur. Sonra bu L model deneme kümesi üzerinde denenerek en başarılı olanı, yani deneme kümesi üzerindeki tahmin başarısı en yüksek olanı seçilir. Eğer bu en iyi model yeterince başarılıysa kullanılır, aksi takdirde başa dönülerek çalışma tekrarlanır [Alpaydın, 2000, s:9].

5.1.4. Birliktelik Kuralları Analizi

Birliktelik Kuralları Analizi veriye uygulanmış, alanlar arasındaki ilişkilerin belirlenebilmesi için ilk olarak SPSS Clementine Programının Web Grafiği aracı kullanılarak mal grupları arasındaki ilişkiler incelenmiştir. Bu araçta 41 mal grubu analize dahil edilmiştir. Mal grupları arasındaki bağıntıları gösteren Web Grafiği ise Tablo 8'de verilmiştir.

TABLO 8: Gruplanmış Mal Grupları Arasındaki Bağntılar

Tablo 8’de verilen grafik tüm bağlantıları göstermektedir. Aralarındaki bağlantıların yüksek olduğu mal gruplarının belirlenmesi için yine aynı araç kullanılarak bağlantı sayısı artırılmış ve Tablo 9’da görülen grafik elde edilmiştir. En sık birlikte alınan mal grupları Kravat/Papyon gruplanmış mal grubu ile Gömlek mal grubudur. Daha sonrasında bu bağlantıyı Triko/T-Shirt/Büstiye – Pantolon/Şort gruplanmış mal grupları ile Triko-Pantolon/Şort gruplanmış mal grupları arasındaki bağlantılar izlemektedir.

TABLO 9: Sıklığı Yüksek olan Bağntılar

Apriori algoritması kullanılarak tespit edilen birliktelikler ve sepet analizi sonuçları aşağıda belirtilmiştir. Bu sonuçlara göre Tablo 10 incelendiğinde;

Erkek Takım Elbisesi ile Kravat veya Papyon alanların %78.846’sı Gömlek de satın almaktadır. Ancak bu kişiler, toplam müşterilerin yalnızca %1.212’sini oluşturmaktadır.

Gömlek alanların %34.094’ü Kravat ya da Papyon da satın almaktadır ve bu kişiler, toplam müşterilerin %17.618’ini oluşturmaktadır.

Ceket ile Pantolon ve Şort alanların %35.424'ü gömlek de satın almaktadır. Bu kişiler, toplam müşterilerin %1.579'unu oluşturmaktadır.

Aynı zamanda, Bluz ve Tunik alanların %34.856'sı T-shirt/Body/Bustiyer de satın almaktadır ve bu kişiler, toplam müşterilerin %6.887'sini oluşturmaktadır.

TABLO 10: Birliktelik Kuralları Analizi Sonucu

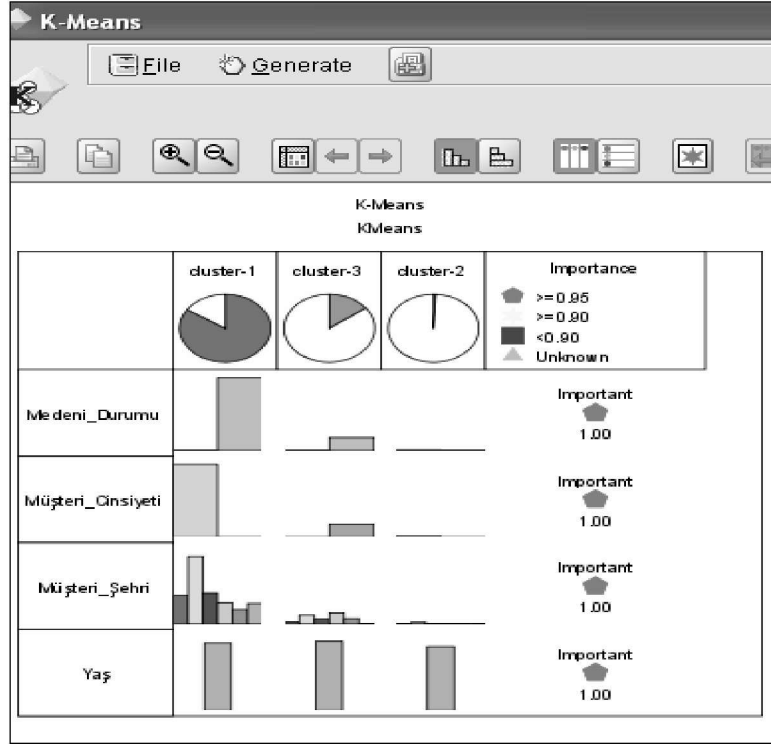
Consequent	Antecedent	Support %	Confidence %
Marka_Mal_Grubu2_Pantolon_Şort	Marka_Mal_Grubu2_TRIKO	1.206	37.681
Marka_Mal_Grubu2_BLUZ/TUNIK	Marka_Mal_Grubu2_ETEK/P.ETEK	1.037	37.079
Marka_Mal_Grubu2_T-SHIRT/BODY/BUSTIER	Marka_Mal_Grubu2_Pantolon_Şort	1.579	36.162
Marka_Mal_Grubu2_T-SHIRT/BODY/BUSTIER	Marka_Mal_Grubu2_Jean	2.272	35.897
Marka_Mal_Grubu2_TRIKO	Marka_Mal_Grubu2_SWEATSHIRT	3.502	35.774
Marka_Mal_Grubu2_TRIKO	Marka_Mal_Grubu2_KAŞKOL	1.06	35.714
Marka_Mal_Grubu2_Pantolon_Şort	Marka_Mal_Grubu2_CEKET	4.434	35.611
Marka_Mal_Grubu2_GÖMLEK	Marka_Mal_Grubu2_CEKET	1.579	35.424
Marka_Mal_Grubu2_GÖMLEK	Marka_Mal_Grubu2_Pantolon_Şort	3.979	35.139
Marka_Mal_Grubu2_Pantolon_Şort	Marka_Mal_Grubu2_ETEK/P.ETEK	1.095	35.106
Marka_Mal_Grubu2_T-SHIRT/BODY/BUSTIER	Marka_Mal_Grubu2_BLUZ/TUNIK	6.887	34.856
Marka_Mal_Grubu2_Kravat_Papyon	Marka_Mal_Grubu2_GÖMLEK	17.618	34.094
Marka_Mal_Grubu2_TRIKO	Marka_Mal_Grubu2_ETEK/P.ETEK	4.83	33.655
Marka_Mal_Grubu2_TRIKO	Marka_Mal_Grubu2_GÖMLEK	1.806	33.226

5.1.5. Kümeleme Analizi

Çalışmanın bu bölümünde, Kümeleme Analizi veri setine uygulanmış ve analiz sonuçları sunulmuştur. Kümeleme modellerinde amaç, küme üyelerinin birbirlerine çok benzediği, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere bölünmesidir[Akpınar, 2000, s:5]. Uygulamada, K-Ortalamlar (K-means) algoritması kullanılmıştır.

5.1.6. K-Ortalamlar Yöntemi

Bu aşamada K-Ortalamlar Yöntemi veri seti üzerinde uygulanmıştır. Analiz sonucunda üç küme elde edilmiştir. SPSS Clementine programı tarafından önem derecesi %90'dan büyük olarak belirlenen değişkenler açıklanmıştır. Elde edilen üç kümede hangi alanların etkili olduğu Tablo 11'de görülmektedir.

TABLO 11: K-Ortalamalar Yöntemi ile Elde Edilen Kümeler

Birinci kümede Medeni Durum, Müşteri Cinsiyeti ve Müşteri Şehri değişkenleri önemli olan değişkenlerdir. Medeni duruma bakıldığında evlilerin, müşteri cinsiyetine bakıldığında ise erkeklerin çoğunlukta olduğu görülmektedir. Müşteri şehri incelendiğinde, İstanbul Avrupa yakasından olan müşterilerin çoğunlukta olduğu görülmektedir.

İkinci kümede önemli olan değişkenler Medeni Durum, Müşteri Cinsiyeti ve Müşteri Şehri'dir. Medeni duruma bakıldığında bekarların, müşteri cinsiyetine bakıldığında ise erkeklerin çoğunlukta olduğu görülmektedir. Müşteri şehri incelendiğinde, İstanbul Avrupa yakasından olan müşterilerin çoğunlukta olduğu görülmektedir.

Üçüncü kümede önemli olan değişkenler Medeni Durum, Müşteri Cinsiyeti, Müşteri Şehri ve Yaş'tır. Medeni duruma bakıldığında evlilerin, müşteri cinsiyetine bakıldığında ise kadınların çoğunlukta olduğu görülmektedir. Müşteri şehri incelendiğinde, Ankara müşterilerinin çoğunlukta olduğu görülmektedir. Yaş ortalaması 45 olan müşteriler bu kümede yer almaktadır.

Değişkenler tek tek ele alınarak incelendiğinde ise şu sonuçlar elde edilmiştir:

Medeni Durum değişkeni incelendiğinde, bu değişkenin her üç küme için de önemli olduğu görülmektedir. Birinci ve üçüncü kümede evli, ikinci kümede ise bekar müşteriler çoğunlukta.

Müşteri Cinsiyeti değişkeni incelendiğinde, bu değişkenin her üç küme için de önemli olduğu görülmektedir. Birinci ve ikinci kümede erkek, üçüncü kümede ise kadın müşterilerin çoğunlukta olduğu görülmektedir.

Müşteri Şehri değişkeni incelendiğinde, bu değişkenin her üç küme için de önemli olduğu görülmektedir. Birinci ve ikinci kümede müşterilerin çoğunluğunun İstanbul Avrupa yakasından, üçüncü kümedeki müşterilerin çoğunluğunun ise Ankara'dan olduğu görülmektedir.

Yaş değişkeni yalnızca üçüncü kümede ayırıcı bir özelliğe sahiptir. Bu kümede yaş ortalaması 45'tir.

Tablo 12'de üç küme dikkate alındığında Birliktelik analizi sonucu hedef değişken olarak belirlenen Gömlek_Kravat_Papyon grubuna üye kişilerin birinci kümede daha büyük oranda oldukları görülmektedir.

TABLO 12: Kümelere Ait Dağılım

Value	Proportion	%	Count
cluster-1		84.25	14461
cluster-2		0.84	144
cluster-3		14.91	2559

\$R-Gömlek_Kravat_Papyon

T

SONUÇ

Bu çalışmada hazır giyim perakende sektöründe faaliyet gösteren bir firmanın alışveriş kayıtları ile alışverişi gerçekleştiren müşteri verileri üzerinde Veri Madenciliği teknikleri kullanılarak bir uygulama gerçekleştirilmiştir. Firmanın verilerinden anlamlı bilgiler çıkarılması için yapılan analizler için SPSS Clementine programı kullanılmıştır. Çalışmada öncelikle müşterilerin demografik özelliklerine göre profilleri belirlenmiş ve demografik özellikleri ifade eden değişkenlerin istatistikleri incelenmiştir. Değişkenlere ait grafikler ile değişkenlerin ortalamaları, minimum ve maksimum değerleri sunulmuştur.

Birliktelik Kuralları Analizi ile müşterilerin alışveriş alışkanlıkları belirlenmeye çalışılmıştır. Yapılan analizler sonucunda gömlek alan müşterilerin %34.094'ünün Kravat ya da Papyon da satın aldığı ve bu kişilerin, toplam müşterilerin %17.618'ini oluşturduğu görülmektedir.

Veri setini oluşturan müşterilerin yaş ortalamasınının 43, cinsiyete göre dağılıma bakıldığında müşterilerin çoğunluğunun erkek olduğu görülmüştür. Yapılan alışverişlerde alınan ürünlerin çoğunluğu kadın giyimidir. Müşterilerin çoğunluğu evlidir ve İstanbul Avrupa yakasında ikamet etmektedir.

Çalışmamızda **Kümeleme Analizi** tekniklerinden **K-Ortalamlar** kullanılmış, bu yöntem ile müşteriler demografik özelliklerine göre kümeler ayrılmıştır. Çalışmada müşterilerin demografik özelliklerinin elde edilen kümeler üzerindeki etkileri k-ortalamlar ile ayrı ayrı incelenmiştir. K-Ortalamlar tekniği kullanılarak oluşturulan üç küme ile kümeleme analizi yapıldığında, hedef değişken olarak belirlenen Gömlek_Kravat_Papyon grubuna üye kişilerin birinci kümede daha büyük oranda olduğu görülmektedir. Birinci kümede Medeni Durum, Müşteri Cinsiyeti ve Müşteri Şehri değişkenleri önemli olan değişkenlerdir. Medeni duruma bakıldığında evlilerin, müşteri cinsiyetine bakıldığında ise erkeklerin çoğunlukta olduğu görülmektedir. Müşteri şehri incelendiğinde ise, İstanbul Avrupa yakasından olan müşterilerin çoğunlukta olduğu görülmektedir. Uygulanan “Birliktelik” ve “Kümeleme” analizleri yardımı ile elde edilen sonuçlardan yararlanılarak, işletme için uygun bir pazarlama stratejisi geliştirilebilir.

Günümüzde, ticari rekabette üstünlük sağlamak için, veriler ayrıntılı olarak kayıt sisteminde yer almalıdır. Ayrıntılı olarak tutulacak bu kayıtların incelenmesi ile, veriler arasında gizlenmiş olabilecek ve yeni fırsatlar doğuracak yapılar ve kurallar belirlenebilecektir. Bu nitelikler işletmeye gerek satış, reklam ve promosyon faaliyetlerinde yol gösterici olabileceği gibi, üretimin satış temelli olarak düzenlenmesine de yardımcı olacaktır. Yığın veriler üzerinden yapılacak Veri Madenciliği çalışmaları, özellikle müşteri odaklı pazarlarda faaliyet gösteren işletmelerin, rekabet üstünlüğü sağlamak üzere pazarlama stratejilerini geliştirmede yardımcı olabilecek bir karar destek aracıdır.

KAYNAKÇA

AAKER, D., 1971, **Multivariate Analysis in Marketing: Theory And Application**, Wadsworth Publishing, California, 299.

AKPINAR, H., 2000, “Veri tabanlarında Bilgi Keşfi ve Veri Madenciliği”, **İ.Ü.İşletme Fakültesi Dergisi**, 29(1), s. 1-22.

ALPAYDIN, E., Z., 2000, “Veri Madenciliği – Ham Veriden Altın Bilgiye Ulaşma Yöntemleri”, **Bilişim 2000 Eğitim Semineri**, 9.

BARBER, F.A., Thomas, R.K., 2001, Huang M.C., “Developing a Profile of LASIK Surgery Customers”, **Marketing Health Services**. 21 (2), s: 32-36.

BERKHIN, P., 2002, **Survey of Clustering Data Mining Techniques**, (Çevrimiçi) http://www.ee.ucr.edu/~barth/EE242/clustering_survey.pdf, California, U.S.A.

BRIJS, T., Swinnen, G., Vanhoof, K., Wets, G., 2004, “Building an Association Rules Framework to Improve Product Assortment Decisions”, **Data Mining and Knowledge Discovery**, 8(1), s: 7-23 .

DECKER, Reinhold, K. Monien, 2003, “Market Basket Analysis with Neural Gas Networks and Self-Organising Maps”, **Journal of Targeting, Measurement and Analysis for Marketing**, 11, s: 373–386.

FAYYAD, U., Piatetsky-Shapiro, G., Smyth, P., 1996, “From Data Mining to Knowledge Discovery in Databases”, **AI Magazine**, 17 (3), s: 37 .

HAIR, J.F., Black, W., Babin, T., Anderson, R., 1998, **Multivariate Data Analysis**, Prentice Hall, U.S.A., s: 473.

KOÇ, S., “Türkiye’de İllerin Sosyo-Ekonomik Özelliklere Göre Sınıflandırılması”, **Çukurova Üniversitesi Sempozyum Bildirisi**, <http://idari.cu.edu.tr/sempozyum/bil20.htm>, 27.06.2005.

NAKİP, M., 2003, **Pazarlama Araştırmaları**, Seçkin Yayınevi, Ankara, S: 418

ÖZEKEŞ, S., 2003, “Data Mining Models and Application Areas”, **Istanbul Commerce University Journal of Science**, 3, s: 65-82.

ÖZER, M., 2001, “User Segmentation of Online Music Services Using Fuzzy Clustering”, **Omega**, 29 (2), s: 193.

ÖZMEN, Ş., 2001, “İş Hayatı Veri Madenciliği Uygulamaları ile İstatistik Uygulamalarını Yeniden Keşfediyor”, **V. Ulusal Ekonometri ve İstatistik Sempozyumu**, Çukurova Üniversitesi, Adana, s: 1- 6

PANDE, J.K., *Analysis of Data Mining Algorithms*,
http://userpages.umbc.edu/%7Ekjoshi1/data-mine/proj_rpt.htm#assoc_rules, 01.01.2007.

PİRAMUTHU, S., 2004, “*Evaluating Feature Selection Methods for Learning in Data Mining Applications*”, *European Journal of Operational Research*, 156 (2), s: 483-494.

RAKESH, A., 1993, “Mining Association Rules Between Sets of Items in Large Databases”, *ACM SIGMOD Konferansı Bildirisi*, U.S.A., s.2.

RAKESH, A., Ramakrishnan, S., 1994, “*Fast Algorithms for Mining Association Rules*”, *20. VLDB Konferansı*, Şili, s: 487-499.

SAİT, R., “*Perakendeciliğin Tanımı*”, <http://www.kobifinans.com.tr/index.php>, 31.12.2005.

SAS Institute Inc., 2009, *The Data Mining Challenge: Turning Raw Data Into Business Gold*, www.sas.com/software/data_mining.

SİLAHTAROĞLU, G., 2008, *Kavram ve Algoritmalarıyla Temel Veri Madenciliği*, Papatya Yayıncılık, İstanbul.

TAN, P.N., Steinbach, M., Kumar, V., 2005. *Introduction to Data Mining*, Addison Wesley, s.332.