

KÜMELEME ANALİZİ VE PAZARLAMADA KULLANIMI

Şebnem KOLTAN YILMAZ*
Said PATİR**

Özet

Kümeleme analizi, veritabanlarındaki verilerin gruplar veya kümeler altında toplanarak, benzer özelliklere sahip nesnelerin bir araya gelmesini sağlayan bir veri madenciliği tekniğidir. Bu çalışmada amaç, pazarlama alanında işletmelerin dikkate alması gereken en önemli ölçütlerin saptanmasında kümeleme analizi tekniğinin uygun bir yöntem olarak kullanılabilceğini göstermektir. Çalışmada, yeni bir halı temizleyicisini pazarlayan bir işletmenin tüketici satın alma tercihleri üzerinde etkili 5 faktöre göre oluşturulan 22 profilden yararlanarak kümeleme analizi tekniği uygulanmıştır. Bu amaçla “Hiyerarşik Kluster (küme) Analizi” kullanılmıştır. Analiz sonuçları, Dendrogram ve Aglomeratif Çizelge ile de desteklenmiştir. Analizde mesafe olarak kullanılan Kareli Öklid ve Pearson Yakınlık Matriksi ölçütlerine göre de birbirlerine en çok benzeyen profiller ve en az benzerlik gösteren profiller saptanmıştır.

Anahtar Kelimeler: Kümeleme Analizi, Pazarlama, Aglomeratif Çizelge, Dendrogram

CLUSTER ANALYSIS AND ITS USAGE IN MARKETING

Abstract

Cluster analysis is one of the data mining techniques which identifies groups of samples that behave similarly or show similar characteristics. The aim of the study indicate cluster analysis techniques are available method that determining the most important factors which businesses in marketing must taken into consideration for evaluating their performances. Cluster analysis technique is applied on 22 buying preference profiles about 10 consumers and 5 factors for a firm that marketing for a new carpet cleaner. For this aim, used hierarchial cluster analysis. The analyses were supported with Dendrogram and Agglomerative Schedule. According to results of Squared Oklid Distance Calculations and Pearson Proximity Matrix, the highest resemblance profiles and the least resemblance profiles are determined.

Keywords: Cluster Analysis, Marketing, Agglomerative Schedule, Dendrogram.

1. GİRİŞ

Ham veri kendi başına değersizdir. Veri, bilgisayar sistemleriyle belirli bir amaca yönelik işlenerek bilgiye dönüşmektedir. Birbiriyle ilişkili verilerin bir arada toplanmasıyla

* Öğr. Gör., İnönü Üniversitesi, Akçadağ MYO, Bilgisayar Programcılığı Bölümü.

**Doç. Dr., İnönü Üniversitesi, İİBF, İşletme Bölümü.

bilgiye erişim ihtiyacı doğmuştur. Organizasyonlarda bu amaca yönelik bilgi sistemleri kurulmaktadır. Bu bilgi sistemleri, satışlar, faturalar, ödemeler gibi işlemlerin takip edilmesini sağlarken, karar almaya yönelik ihtiyaçlara da girdi oluşturmaktadır. Ancak bu sistemler karar almada kolaylık sağlarken bunu, önceden belirlenmiş ve sistem içerisinde yapılandırılmış konularda gerçekleştirebilmektedir. Kurum içi ve dışı bilgilerin yanı sıra önceden tahmin edilemeyen sorulara yönelik cevap bulan, ileriye dönük tahmin sistemlerine ihtiyaç duyulmuştur. Bu bağlamda veri madenciliği (VM) teknikleri diğer birçok alanda olduğu gibi işletmecilik alanında da yaygın bir şekilde kullanılmaktadır. VM, bilgi teknolojilerinin doğal gelişim sürecinin sonucu olarak da değerlendirilebilir.

VM amaçlarından biri olan kümeleme, istatistiksel veri analizi, örüntü tanıma gibi birçok alanda oldukça sık kullanılmaktadır. Veritabanlarındaki verilerin gruplar veya kümeler altında toplanarak, benzer özelliklere sahip nesnelerin bir araya gelmesini sağlayan kümeleme algoritmaları veri madenciliği alanında oldukça büyük öneme sahiptir.

Veri kümeleme güçlü bir gelişme göstermektedir. Veri tabanlarında toplanan veri miktarının artmasıyla orantılı olarak, kümeleme analizi son zamanlarda veri madenciliği araştırmalarında aktif bir konu haline gelmiştir.

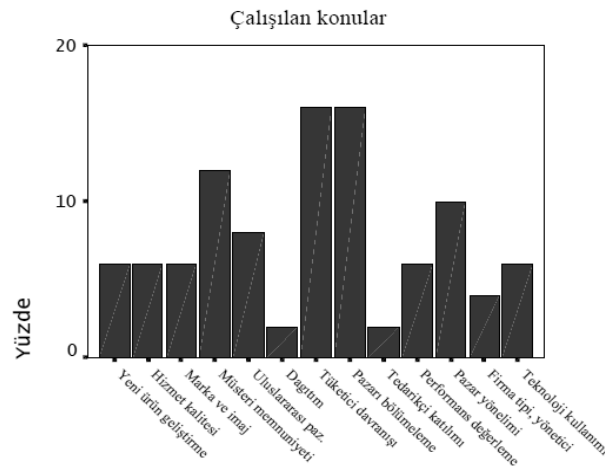
Green vd.(1967), sayısal sınıflamanın gerekli olduğu birçok pazarlama probleminin çözümünde kümeleme analizinin faydalı olacağını ifade etmektedirler. Araştırmada, kümeleme analizinin, çeşitli malların, dağıtım kanallarının ve tüketici gruplarının bir arada analizinde kullanılabilen bir yöntem olduğu belirtilerek; piyasa yapısı analizlerinde ve bir piyasanın diğer piyasalarla benzerlik veya farklılıklarının karşılaştırılması gerektiğinde kümeleme analizinin çok sayıda değişkeni dikkate alması nedeniyle kullanışlı bir teknik olduğu sonucuna varılmaktadır. Day ve Heeler (1971), kümeleme analizinin, yeni ürünlerin pazar denemelerini arttırmak ve geliştirmek için, homojen pazar bölümlendirmelerini tanımlayabilmek amacıyla kullanılabileceğini göstermektedirler. Klasterin (1983), çalışmasında kümeleme analizi sonuçlarının istatistik değerlendirmesini yapmıştır. Harrigan (1985), çalışmasında aynı pazarda yer alan rakip işletmeler için faydalı olan stratejik gruplar geliştirmek amacıyla kümeleme analizini kullanmıştır. Sackson (1990), çalışmasında işletme politikaları ve stratejilerinin bir parçası olan işletme planlarını geliştirmek amacıyla kümeleme analizini kullanmıştır. Ketchen and Shook (1996), çalışmasında stratejik yönetim

alanında bir kümeleme analizi örneği göstermiştir. Yüksel (2002), dondurulmuş hazır yemek satın alan tüketicilerin kendi aralarında belirli özelliklere sahip segmentlerde toplanıp toplanmadıklarını incelediği tanımlayıcı nitelikteki araştırmasında, kümeleme analizi sonucunda, dondurulmuş hazır yemek satın alıp kullananların, satın alma nedenlerine göre anlamlı olarak dört segment oluşturdukları sonucuna ulaşmıştır. Arslan (2004), çalışmasında, etkin bir spor ayakkabısı pazarı oluşturabilmek için tüketici bölümlerinin oluşturulmasında çok boyutlu bir yaklaşım önerisi önermiştir. Bu amaçla oluşturduğu verilere kümeleme analize ve ayırma analizi uygulamış, bu sayede tüketici bölümleri oluşturmuştur. Demiralay ve Çamurcu (2005), çalışmada, hiyerarşik kümeleme algoritmalarından CURE (Clustering Using REpresentatives) ve AGNES (AGglomerative NEsting) ile bölümleyici kümeleme algoritmalarından çok sık kullanılan K-Means' in sentetik veri setlerinde uygulanmasıyla elde edilen sonuçların karşılaştırmışlardır. Aşan (2007), çalışmasında kümeleme analizi kullanarak kredi kartı kullanan banka müşterilerinin sosyo-ekonomik özelliklerini gruplamıştır. Canitez ve Yeniçeri (2007), çalışmasında, ihracat yapan KOBİ'lerin ihracat performansı itibarıyla, pazar yönlülük bakımından farklarını belirlemeye yönelik yapılan bu çalışmada, araştırma hipotezlerini test etmeye geçmeden önce KOBİ'leri gruplandırabilmek amacıyla hiyerarşik olmayan kümeleme analizi uygulamıştır. Atbaş (2008), kümeleme analizinde küme sayısının belirlenmesi üzerine bir çalışma yapmıştır. Doğan (2008), çalışmasında Türk Bankacılık Sektörü'nde (1998–2006) dönemi itibarıyla faal olan ticaret bankalarına ait finansal oranlar temel alınarak yapılan “Kümeleme Analizi” uygulamaları ile elde edilen sonuçlar ve bankalar için yapılan finansal analiz sonuçları ile uyumluluğu tartışılarak elde edilen sonuçlar ışığında “Kümeleme Analizi” tekniğinin bankaların finansal performanslarını belirlemek ve finansal açıdan benzer bankaları tanımlamak amacıyla, bankaların gözetiminde kullanılan mevcut teknikleri tamamlayıcı bir teknik olarak kullanılabilirliği irdelenmektedir. Ergün (2008), çalışmasında ürün kategorileri ve ürün sınıfları arasındaki satış ilişkisinin tespit edilmesini amaçlamıştır. Bu amaçla perakendeci bir işletmenin bir yıllık süre boyunca topladığı alışveriş fişi verileri üzerinde birliktelik kuralları analizi ve hiyerarşik kümeleme gibi veri madenciliği yöntemleri uygulamıştır. Yücel ve Yeniçeri (2008), ihracat yapan küçük ve orta büyüklükteki işletmelerin ihracat performansı, insan sermayesi, örgütsel sermaye ve bakımından benzerlik gösterip göstermediğini belirlemeye çalışmışlardır. Araştırma hipotezini test edebilmek amacıyla hiyerarşik olmayan kümeleme analizi kullanılmıştır. Helgesen vd. (2009), çalışmasında pazarlama alanındaki tecrübesiz elemanların algıları ve iş

performansları arasındaki ilişkiyi analiz etmeyi amaçlamışlardır. Bu amaçla kümeleme analizini kullanmışlardır. Funfgeld ve Wang (2009), çalışmasında, günlük mali işlerdeki davranışlar ve tutumlar üzerinde kümeleme analizi uygulamışlardır. Mehrizi ve Bontis (2009), bilgi yönetimi (BT) alanında literatür taraması yaparak bunların içeriklerini kümeleme analizi ile ortaya koymuştur.

Çetin (2003), 1995–2002 yılları arası pazarlama araştırmalarında, çok değişkenli istatistiksel analizlerin kullanımının genel bir değerlendirmesini yapmak amacıyla bir çalışma hazırlanmıştır. Bu araştırmanın sonuçları şöyledir:

Grafik 1.1: İncelenen Makalelerde Çalışılan Konular

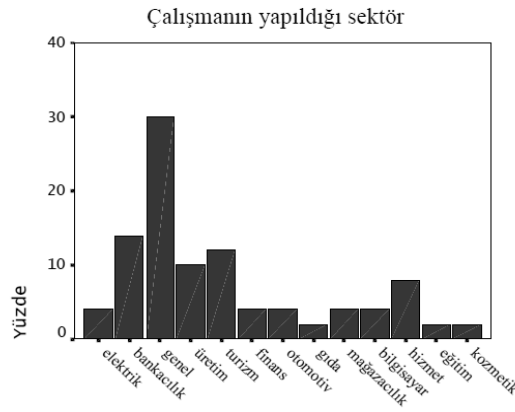


Kaynak: Çetin, Emre İpekçi (2003), Çok Değişkenli Analizlerin Pazarlama İle İlgili Araştırmalarda Kullanımı: 1995–2002 Arası Yazın Taraması Akdeniz İ.İ.B.F. Dergisi (5) s.39.

Grafik 1.1’de pazarlama ile ilgili araştırmalarda çalışılan konular açısından bir değerlendirme yapıldığında, pazar bölünmeleme (%16) ile tüketici davranışları (%16) konusunda yoğunlaştığı ve bu konuları tüketici memnuniyeti (%12) ve pazar yönelimi (%10) konularının takip ettiği görülmektedir.

İncelenen 50 makale uygulamaların yapıldığı sektörler açısından değerlendirildiğinde (Grafik 1.2) ise genel nitelikli çalışmaların (%30) yanı sıra, bankacılık (%14), turizm (%12) ve üretim (%10) sektörlerinin öne çıktığı görülmektedir.

Grafik 1.2: İncelenen Makalelerde Çalışılan Sektörler



Kaynak: Çetin, Emre İpekçi (2003), Çok Değişkenli Analizlerin Pazarlama İle İlgili Araştırmalarda Kullanımı: 1995–2002 Arası Yazın Taraması Akdeniz İ.İ.B.F. Dergisi (5) s.40.

Pazarlama araştırmaları konularını ele alan 50 bilimsel makalede faydalanılan çok değişkenli istatistiksel analizler ve bu analiz yöntemlerinin birlikte kullanımı değerlendirilerek Tablo 1’de özetlenmiştir. Tablo 1 incelendiğinde, çalışmaların 33’ünde (%66) faktör analizinin kullanıldığı görülmüştür. Bu çalışmaların %24’ünde sadece faktör analizi kullanılmışken, çalışmaların %16’sında faktör analizinin kümeleme analizi ile %10’unda diskriminant analizi ile %12’sinde anova-manova analizi ile %16’sında çoklu regresyon analizi ile ve %2’sinde correspondence analizi ve çok boyutlu ölçekleme analizi ile birlikte kullanıldığı görülmüştür.

Tablo 1.1: İncelenen Makalelerde Çok Değişkenli Analizlerin Birlikte Kullanım Oranları

| | | Faktör Analizi | Kümeleme Analizi | Diskriminant Analizi | Anova-Manova Analizi | Çoklu Regresyon Analizi | Correspondence Analizi | Conjoint Analizi | Çok boyutlu ölçekleme | Toplam (*) |
|-------------------------|------------|----------------|------------------|----------------------|----------------------|-------------------------|------------------------|------------------|-----------------------|------------|
| Faktör Analizi | Yüzde Sayı | %24 12 | %16 8 | %10 5 | %12 6 | %16 8 | %2 1 | - | %2 1 | %66 33 |
| Kümeleme Analizi | Yüzde Sayı | | %2 1 | %14 7 | %10 5 | %2 1 | - | %2 1 | %4 2 | %28 14 |
| Diskriminant Analizi | Yüzde Sayı | | | %0 0 | %6 3 | %4 2 | - | %2 1 | - | %20 10 |
| Anova-Manova Analizi | Yüzde Sayı | | | | %6 3 | %8 4 | - | %4 2 | %2 1 | %28 14 |
| Çoklu Regresyon Analizi | Yüzde Sayı | | | | | %6 3 | - | %2 1 | - | %30 15 |
| Correspondence Analizi | Yüzde Sayı | | | | | | %2 1 | - | - | %2 1 |
| Conjoint Analizi | Yüzde Sayı | | | | | | | %4 2 | - | %4 2 |
| Çok Boyutlu Ölçekleme | Yüzde Sayı | | | | | | | | %4 2 | %4 2 |

(*) Toplamda yer alan değerler ile aynı satırlara ait değerlerin toplamının aynı olmamasının nedeni bazı makalelerde 2'den fazla analiz kullanılmış olmasından kaynaklanmaktadır.

Kaynak: Çetin, Emre İpekçi (2003), Çok Değişkenli Analizlerin Pazarlama İle İlgili Araştırmalarda Kullanımı: 1995–2002 Arası Yazın Taraması Akdeniz İ.İ.B.F. Dergisi (5) s.41.

Çalışmaların %28'inde kümeleme analizinden faydalandığı tespit edilmiştir. Kümeleme analizinin tek başına kullanıldığı makale oranı %2 iken, diğer çok değişkenli istatistiksel yöntemlerle bir arada kullanımı incelendiğinde, çalışmaların %14'ünde kümeleme analizinin diskriminant analizi ile %10'unda anova-manova analizi ile %4'ünde çok boyutlu ölçekleme analizi ile ve %2'sinde çoklu regresyon ve conjoint analizi ile birlikte kullanıldığı ortaya çıkmıştır (kümeleme analizi ile faktör analizinin bir arada kullanım oranı daha önce de belirtildiği gibi %16'lık kısmı oluşturmaktadır).

2. VERİ MADENCİLİĞİ

Dijital veri elde etme ve depolama teknolojisindeki gelişmeler, çok büyük miktarlardaki veri tabanlarının artmasıyla sonuçlanmıştır. Bu artış süpermarkete ait veriler, kredi kartı kullanma kayıtları, telefon arama detayları ve devlet istatistikleri gibi günlük, olağan olaylardan astronomik miktarlardaki görüntüler, moleküler veri tabanları ve tıbbi kayıtlar gibi farklı olaylara kadar insan çabalarının her alanında meydana gelmiştir (Hand vd., 2001:3). Çok büyük ölçekli veriler, farklı alanlardaki büyük ölçekli veri tabanları içinde değerli verileri bulduran bir veri madeni gibi düşünülebilir. VM ise bu verilerden önceden bilinmeyen anlamlı bilgileri üretme süreci olarak tanımlanmaktadır. Diğer bir anlatımla VM,

veri tabanlarında yer alan büyük hacimli verilerde bulunan gizli bilgilerin, örüntülerin, kuralların ve ilişkilerin gelişmiş çok boyutlu tekniklerle keşfedilmesi süreci olarak tanımlanabilir. VM, bu süreci bilgisayar, makine öğrenmesini (machine learning), veri tabanı veya veri ambarı yönetimini, matematiksel algoritmaları ve istatistik teknikleri kullanarak gerçekleştirmektedir.(Yılmaz, 2008:1).

VM genel anlamda; büyük miktarda veri içerisinde, gizli kalmış, değerli, kullanılabilir bilgilerin açığa çıkarılması biçiminde ifade edilmektedir. Bu ifade irdelendiğinde yakın zamana kadar pek duyulmamış bir terim olan VM' nin, aslında uzun zamandır gerçekleştirilmesine karşın evrim geçirerek 1990'lardan itibaren isimlendirildiği görülmektedir. Öyle ki, bu konu en iyi biçimde 'istatistikçiler, VM'ni elle yaparlar,' ifadesiyle özetlenebilir. Bu ifade her ne kadar VM' nin temel mantığının klasik istatistik teknikler olduğunun altını çizse de, herhangi bir istatistik yazılımı veya yöntemini VM olarak ifade etmek doğru olmamaktadır. Bununla beraber VM' yi, istatistik tekniklerin evrim geçirmiş hali olarak düşünmenin de mümkün olmasının yanı sıra; VM' nin, kendisine özgü tanımlayıcı özellikleri vardır (Koyuncugil ve Özgülbaş, 2006:4).

VM' de kullanılan modeller, tahmin edici (predictive) ve tanımlayıcı (descriptive) olmak üzere iki ana başlık altında incelenmektedir. VM modellerini gördükleri işlemlere göre sınıflama (classification) ve regresyon (regression) modelleri, kümeleme (clustering) modelleri ve birliktelik kuralları (association rules) ve ardışık zamanlı örüntüler (sequential patterns) olmak üzere üç ana başlık altında incelemek de mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir. (Albayrak ve Yılmaz, 2009:33).

VM amaçlarından biri olan kümeleme, istatistiksel veri analizi, örüntü tanıma vb, birçok alanda oldukça sık kullanılmaktadır. Veritabanlarındaki verilerin gruplar veya kümeler altında toplanarak, benzer özelliklere sahip nesnelerin bir araya gelmesini sağlayan kümeleme algoritmaları veri madenciliği alanında oldukça büyük öneme sahiptir. Ancak kümeleme algoritmalarının seçimi hem mevcut veri türüne hem de amaç ve uygulamaya bağlıdır. Şayet kümeleme analizi tanımlayıcı ya da açıklayıcı bir araç olarak kullanılacaksa bu durumda aynı veri seti üzerinde birkaç algoritmanın uygulanması mümkün olmaktadır. Bunlar arasında veritabanı teknolojisine katkıda bulunan algoritmalar olarak, CLARANS, BIRCH, DBSCAN,

CURE, STING, CLIQUE ve WaveCluster algoritmaları sayılabilir. Bütün bu algoritmalar büyük ölçekli veritabanlarında çok büyük miktarda bulunan verilerin kümeleme problemlerini çözmeye çalışmaktadır.(Akın, 2008:2)

Kümeleme analizi sınıflandırma modellerinden farklıdır. Kümeleme analizinde, sınıflama modelinde olan veri sınıfları yoktur. Verilerin herhangi bir sınıfı bulunmamaktadır. Sınıflama modelinde, verilerin sınıfları bilinmekte ve yeni bir veri geldiğinde bu verinin hangi sınıftan olabileceği tahmin edilmektedir. Oysa kümeleme analizinde, sınıfları bulunmayan veriler gruplar halinde kümeler ayrılırlar. Bazı uygulamalarda kümeleme analizi, sınıflama modelinin bir önişlemi gibi görev alabilmektedir (Özekes, 2003:71).

Marketlerde farklı tüketici gruplarının keşfedilmesi ve bu grupların alışveriş örüntülerinin ortaya konması, biyolojide bitki ve hayvan sınıflandırmaları ve işlevlerine göre benzer genlerin sınıflandırılması, şehir planlanmasında evlerin tiplerine, değerlerine ve coğrafik konumlarına göre gruplara ayrılması gibi uygulamalar tipik kümeleme uygulamalarıdır. Kümeleme aynı zamanda Web üzerinde bilgi keşfi için dokümanların sınıflandırılması amacıyla da kullanılabilir (Özekes, 2003:72). Kümeleme analizinin işletme biliminde kullanımının pazarlama araştırmaları alanında yoğunlaştığını ve stratejik grupları tanımlamak, belirli bir mal veya hizmetin hitap ettiği tüketici gruplarını sınıflandırmak, bölgeler arası tüketim benzerliklerini/farklılıklarını belirlemek ve ürünleri tüketici karakteristiklerine göre sınıflandırmak gibi amaçlarla kullanıldığının görüldüğünü belirtmektedir (Doğan, 2008:78).

3. KÜMELEME ANALİZİ

İnsanlar önemli kararlar almadan önce genellikle bir adım geri atar ve büyük resmi görmeye çalışırlar. Ancak zaman zaman bu büyük resim anlaşılacak kadar karmaşıktır. Büyük bir veritabanı çok miktarda boyut, alan içerebilir ve çok karmaşık bir yapıya sahip olduğundan en iyi uygulanabilen VM yöntemleri bile bu veri yığını içerisinde anlamlı sonuçlar üretemeyebilir. Çok karmaşık ve büyük sorunları çözmekte izlenen yöntem genellikle büyük sorunu daha küçük ve tek basına daha rahat çözülebilecek alt sorunlara bölmek ve her bir alt sorunu çözdükten sonra çözümleri birleştirerek sonuca gitmek şeklindedir. Ancak bazı durumlarda veriler öyle dağılmışlardır ki nereden bölüneceğini ve

hangi şekilde alt gruplara ayrılabilceğini kestirmek mümkün değildir. Bu yüzden otomatik küme bulma yöntemleri geliştirilmiştir (Dolgun, 2006:26).

Sınıflamaları hakkında açık bilgi bulunmayan durumlarda, topluluğa ilişkin tahminlerin yapılmasında yararlanılan bir yöntemler grubu olarak tanımlanan kümeleme analizi, araştırmacıya, üzerinde çalışılan herhangi bir veri setindeki benzer(homojen) birey gruplarını bulma, kendi içinde türdeş fakat diğerlerinden farklı olacak biçimde kümelere ayırma olanağı tanıyan birçok değişkenli istatistiksel analiz tekniğidir . Segment analizi ve taksonomi analizi olarak da adlandırılan kümeleme analizinin genel amacı, gruplanmamış verileri benzerliklerine göre sınıflandırmak ve araştırmacıya uygun, ise yarar ve özetleyici bilgiler elde etmede yardımcı olmaktır (Harrigan, 1985:59; Sambamoorthi, 2010:2; Ketchen and Shook, 1996: 441; Doğan, 2008:76).

Küme, birbirine yakın (benzer) nesnelere çok boyutlu uzayda oluşturdukları bulutlar benzetmesi' şeklinde tanımlanabilir. Kümeleme analizi ise; bu kümeleri oluşturma işlemidir. Kümeleme analizi için yapılan başka bir tanım ise; küme adı verilen nispeten homojen grupların iç içe durumların ya da nesnelere sınıflandırılması için kullanılan teknikler sınıfıdır. (Akın, 2008:5). Böylece nesnelere, örneklenen küme özelliklerini iyi yansıtan etkili bir temsil gücüne sahiptir. Sınıflamanın aksine, yeniden tanımlanmış sınıflara dayalı değildir. Kümeleme, bir denetimsiz öğrenme (unsupervised learning) yöntemidir (Aydoğan,2003:88) . Amaç, elemanların birbirlerine çok benzediği, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere (gruplara) bölünmesidir (Durmuş ve İplikçi, 2007).

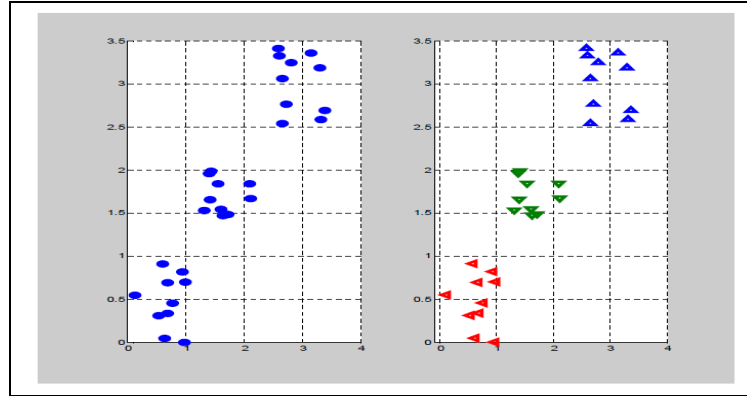
Kümeleme analizi farklı sorunların giderilmesinde kullanılan önemli bir analizdir. Öncelikle, doğru grupların bulunmasında yararlıdır. Örneğin, Avrupa Birliği üye ülkelerin 10 sektördeki istihdam durumlarına göre kümelenebilir. Bu durumda objektif grupları tanımlamak için Kümeleme analizinden yararlanılabilir. İkinci olarak Kümeleme analizi veri indirgemesi içinde yararlıdır. Örneğin şehirlerin çoğu yeni bir ürün için test pazarları olarak kullanılabilir. Eğer şehirler benzer şehirleri içeren küçük sayıdaki gruplara ayrılırsa her bir gruptan bir üye test pazarı için kullanılabilir. Böylece araştırmada gerek zaman gerekse maliyet ve emek açısından önemli tasarruf sağlanmış olur (Akın, 2008:4).

Kümeleme analizi X veri matrisinde yer alan ve doğal grupları kesin olarak bilinmeyen birimleri, değişkenleri ya da birim ve değişkenleri birbirleri ile benzer olan alt kümelere (grup, sınıf) ayırmaya yardımcı olan yöntemler topluluğudur. Kümeleme analizi, birimleri değişkenler arası benzerlik ya da uzaklıklara dayalı olarak hesaplanan bazı ölçülerden yararlanarak homojen gruplar oluşturmaya çalışır (Özdamar, 2004:279).

Kümeleme analizi bağımlı ve bağımsız değişkenler arasında fark ya da üstünlük gözetmez. Tersine birbirine bağımlı tüm ilişkileri inceler. Kümeleme analizi, hedefi belirlenmiş değişkenler grubuna dayanan nispeten homojen grupların içine nesnelere sınıflandırmaktadır. Aynı grup içinde yer alan nesnelere; değişkenler bakımından nispeten birbirlerine benzerlik gösterirken diğer gruptaki nesnelere ise farklıdır (Akın, 2008:6).

Kümeleme analizi; birkaç adımdan oluşan bir çözüm sürecidir. Veri girişi analizin ilk aşamasıdır. Yani ilk olarak doğal sınıflamaları hakkında kesin bilgilerin bulunmadığı ana kütlelerden alınan n sayıda birimin incelenen p sayıda değişkene ilişkin gözlem sonucu değerleri elde edilir. Böylece veri matrisi oluşturulmuş olur. Daha sonra verinin ölçüm tipine uygun bir benzerlik ölçüsü ile nesnelere uzaklıklar matrisi elde edilir. Uygun kümeleme tekniği seçilir ve uygulanır. Tekniğin uygulanması sonucu nesnelere kümelere ayrılmış olur. Kümeleme sonuçlarının anlamlılığının yorumlandığı aşama ise analizin son aşamasıdır. Kümeleme analizi sonucunda kümeleri oluşturan elemanlar birbirine benzerlik, baks kümelerin elemanlarından farklılık gösterirler. Kümeleme işlemi başarılı olursa, bir geometrik çizim yapıldığında birimler küme içerisinde birbirlerine çok yakın, kümeler ise birbirlerinden uzak olacaktır. Şekil 3.1' de basit bir kümeleme işlemi gösterilmiştir (Vatansever, 2008:84).

Şekil 3. 1: Veri Birimleri ve Kümeleme



Kaynak: VATANSEVER, Metin (2008), *Görsel Veri Madenciliği Tekniklerinin Kümeleme Analizlerinde Kullanımı ve Uygulanması*, Yayınlanmamış Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, Fen Bilimler Enstitüsü, İstanbul, s. 85.

Kümeleme (Clustering) işlemi, heterojen yapıya sahip bir kitleyi daha homojen birkaç alt gruba ya da kümeye bölme işlemidir. Sınıflama ile kümelemeyi birbirinden ayıran en önemli fark, kümeleme işleminin sınıflama işleminde olduğu gibi önceden belirlenmiş bir takım sınıflara göre bölme yapmamasıdır. Sınıflamada her bir veri, önceden sınıflandırılmış bir takım sınıflar üzerinde yapılan bir eğitim neticesinde ortaya çıkan bir modele göre, önceden belirlenmiş olan bir sınıfa atanmaktadır. Kümeleme işleminde ise önceden tanımlanmış sınıflar ya da örnek sınıflar bulunmamaktadır. Verilerin kümelenmesi işlemi, verilerin birbirlerine olan benzerliklerine göre yapılmaktadır. Oluşan sınıfların hangi anlamları taşıdığına belirlenmesi tamamen çözümlmeyi yapan kişiye kalmıştır (Vatansever, 2008:53).

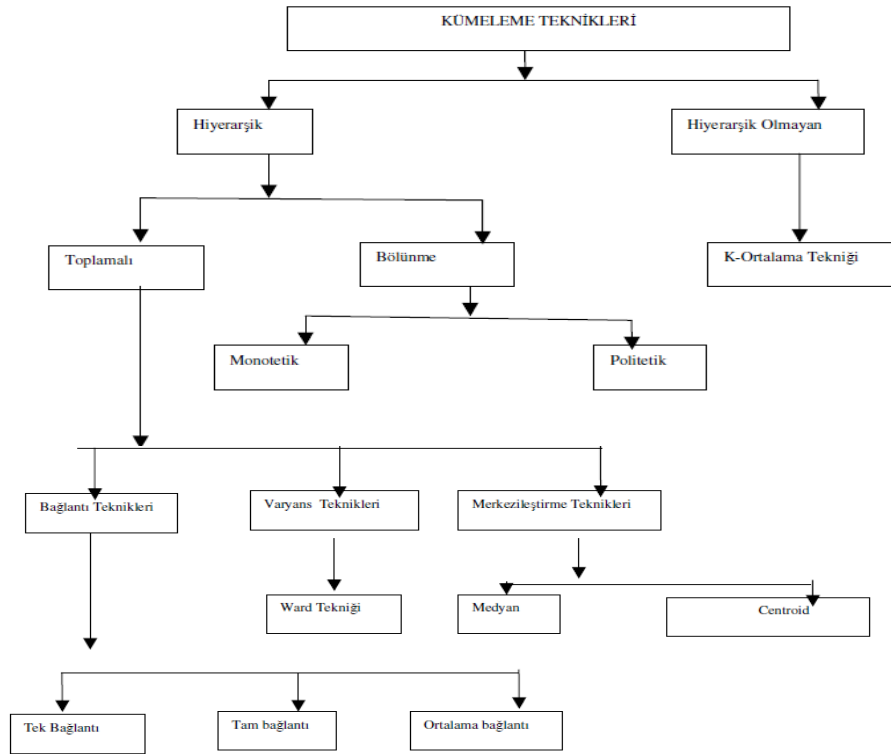
3. 1. Kümeleme Analizi Teknikleri

Kümeleme teknikleri; uzaklık matrisini kullanarak nesnelere veya değişkenlere kendi içinde homojen ve kendi aralarında heterojen gruplar oluşturmaya imkan sağlamaktadır. Kümeleme analizi için birçok algoritma öne sürülmüştür. Ancak literatürde bu algoritmalar iki başlık altında toplanmıştır: Hiyerarşik kümeleme teknikleri, Hiyerarşik olmayan kümeleme tekniği (Ketchen and Shook, 1996: 444). Her iki teknikte de ortak amaç kümeler arasındaki farklılıkları ve kümeler içi benzerlikleri en yüksek düzeye çıkarmaktır. Yani, küme içi homojenlik artırılırken kümeler arası homojenlik ise azaltılmaktadır. Hangi tekniğin kullanılacağı küme sayısına bağlı olmakla birlikte her iki tekniğin beraber kullanılması çok

daha yararlıdır. Böylece hem sonuçları hem de iki tekniğin hangisinin daha uygun sonuçlar verdiğini karşılaştırmak mümkün olmaktadır (Akın, 2008:8).

Kümeleme tekniklerinin genel yapısı Şekil 3.2’de verilmiştir

Şekil 3. 2: Kümeleme Tekniklerinin Genel Görünümü



Kaynak: AKIN, Yasemin Koldere (2008), *Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi*, Yayınlanmamış Doktora Tezi, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul, s. 9.

Hiyerarşik kümeleme yöntemi, kümelerden bir eleman silme ya da eklemeye bir ağaca benzeyen yapı gösteren aşamalar grubudur (Ketchen and Shook, 1996: 444). Aşamalı (hiyerarşik) kümeleme yöntemleri veri setindeki birimlerin birbirilerine göre uzaklık veya benzerliklerini dikkate alarak birimleri birbirleriyle değişik aşamalarda bir araya getirerek ardışık biçimde kümeler belirlemeye ve bu kümelere girecek elemanların hangi uzaklık veya benzerlik düzeyinde küme elemanlarının olduğunu belirlemeye yönelik yöntemlerdir (Özdamar, 2004:293).

Temelde bu algoritmaya dayanan önemli hiyerarşik kümeleme yöntemleri ise şunlardır (Ünlükaplan, 2008:25):

Tek Bağlantı Kümeleme Yöntemi (TEKBKY, SINGLE Linkage, Nearest Neighbor Method): Küme elemanları arasındaki en düşük uzaklık değeri temel alınarak kümelerin oluşturulması esasına dayanır

Ortalama Bağlantı Kümeleme Yöntemi (ORTBKY, AVERAGE Linkage Method): Tek bağlantı ve tam bağlantı tekniği ile paralel olarak, kümeler arasındaki ortalama uzaklığın en düşük değerini temel alır.

Tam Bağlantı Bağlantı Kümeleme Yöntemi (TAMBKY, COMPLETE Linkage Method): İki kümenin elemanları arasındaki en yüksek uzaklığın minimum değerini temel alır.

Küresel Ortalama Bağlantı Kümeleme Yöntemi (KOBKY CENTROID Linkage Method): Bir kümeyi oluşturan gözlemlerin ortalamalarını esas alır. Kümede sadece bir merkez varsa onun değeri merkez olarak kabul edilir.

Ward Bağlantı Bağlantı Kümeleme Yöntemi (WBKY, WARD Linkage Method): Minimum varyans yöntemi olarak da bilinen bu yaklaşım, bir kümenin ortasına düşen gözlemin, aynı kümenin içinde bulunan gözlemlerden ortalama uzaklığını ele alır ve toplam sapma karelerinden yararlanır. Yoğun olarak kullanılan hiyerarşik kümeleme yöntemidir.

Hiyerarşik olmayan kümeleme tekniği ise (aynı zamanda K-Means (K-ortalama) tekniği olarak adlandırılan) veri setlerini önceden belirlenmiş küme gruplarına ayırır (Ketchen and Shook, 1996: 445). Bölmeli kümeleme yöntemleri aşamalı olmayan kümeleme yöntemleridir. Bölmeli yöntemlerde, n birimim $k < n$ olmak üzere k kümeye parçalanması rasgele veya gelişigüzel yapılabilir. Bu yöntemde birimleri ayırmak istediğimiz küme sayısını belirledikten sonra, kümeler için belirlenen küme ayırma kriterlerine göre birimlerin hangi kümelere gireceğine karar verilir ve atama işlemi gerçekleştirilir. Kümeler tarafsız bölme kriteri olarak nitelendirilen bir kritere uygun oluşturulduğu için aynı kümedeki birimler birbirlerine benzerken, farklı kümedeki birimlerden farklıdırlar. Bölmeli yöntemler hiyerarşik yöntemlere oranla daha büyük veri setlerine uygulanabilir (Özdamar, 2004:325).

Yöntemde aşamalı (hiyerarşik) yöntemden farklı olarak küme sayısı araştırmacı tarafından belirlenir. Buna göre, birey ya da nesnelere, kümeler içi kareler toplamı en küçük

olacak biçimde k sayıda kümeye bölünmektedir. Daha sonra her kümenin tipik gözlemi seçilir ve benzer gözlemler, tipik gözlemin etrafında sırayla kümelenir (Ünlükaplan, 2008:25).

2. 2. Uzaklık Ölçüleri

Kümeleme analizinde küme, birbirine yakın nesne ya da bireylerin oluşturdukları grup olarak tanımlanabilir. Kümeleme analizinde birey ya da nesnelere arasındaki uzaklıkları hesaplamak için en yaygın kullanılan uzaklık ölçüsü Öklid uzaklığıdır. Öklid uzaklığı iki obje arasındaki benzerliği ölçmede en yaygın kullanılan uzaklık ölçüsü olup iki obje arasında çizilecek bir düz doğrunun uzunluğunu temel alır. Bu uzaklık ölçüsü dışında aşağıda sunulan uzaklık değerlerinden yararlanılarak birey ya da nesnelere kümelere dahil edilmesi işlemi gerçekleştirilmede kullanılan diğer uzaklık ölçüleri de vardır (Ünlükaplan, 2008:21):

- Minkowski uzaklığı
- Ölçekli Öklid uzaklığı
- Binary Öklid uzaklığı
- Pearson uzaklığı
- Manhattan (City Blok) Uzaklığı
- Mahalanobis Uzaklığı
- Hotelling T2 Uzaklığı
- Canberra Uzaklığı

3. UYGULAMA: SPSS ÖRNEK DOSYASI VERİSİ

3. 1. Çalışmanın Tanıtılması

Bu çalışmada, SPSS 15.0 programına ait Örnek Klasörü (Sample Files) içerisinde bulunan verilerden faydalanılmıştır. Bu amaçla seçilen “carpet. sav” adlı veri dosyasında yer alan değişkenler üzerinde Hiyerarşik Kümeleme Analizi yöntemi kullanılarak tüketici satın alma tercihlerinin benzerlikleri belirlenmiştir.

3. 1. 1. Çalışmanın Amacı ve Önemi

Bu çalışmanın temel amacı, pazarlama işletmelerinin belirli yönlerden performanslarını değerlendirmede dikkate alması gereken en önemli ölçütlerin (değişkenlerin) sınıflandırılmasında kümeleme analizi tekniklerinden hiyerarşik kümeleme tekniğinin uygun bir yöntem olarak kullanılabilceğini göstermektir.

Kümeleme analizi teknikleri, tıp, eğitim, bankacılık, savunma sanayi, coğrafya, sosyal bilimler, çevre bilimleri, ziraat gibi pek çok alanda kullanılmakla birlikte işletme alanında kullanımı da oldukça yaygındır. Literatürde pazarlama alanında ürün tercihleri, pazar bölümlendirme, tüketici satın alma davranışları, dağıtım kanallarının ve tüketici gruplarının analizi, piyasa yapısı analizleri ve bir piyasanın diğer piyasalarla benzerlik veya farklılıklarının karşılaştırılması gibi alanlarda çok sayıda değişkenin dikkate alınması nedeniyle kullanılmaktadır. Pazarlama alanında kümeleme analizinin kullanımıyla ilgili çalışmalar bölüm 1’de ayrıntılı olarak verilmiştir.

3. 1. 2. Çalışmanın Kapsamı

Bu çalışmada, SPSS 15.0 programına ait Örnek Klasörü (Sample Files) içerisinde bulunan verilerden faydalanılmıştır. Bu amaçla seçilen “carpet.sav” adlı veri dosyasında yer alan değişkenler üzerinde Hiyerarşik Kümeleme Analizi yöntemi kullanılarak tüketici satın alma tercihlerinin benzerlikleri belirlenmeye çalışılmaktadır.

İncelen “carpet.sav” adlı dosya şöyledir: (Green and Wind, 1973): Yeni bir halı temizleyicisini pazarlayan bir işletme, tüketici satın alma tercihleri üzerindeki 5 faktörün etkilerini (ambalaj dizaynı, marka ismi, fiyat, Good Housekeeping seal, para iade garantisi) araştırmak istemektedir. Bu amaçla 3 adet ambalaj dizaynı (her biri farklı resimli), 3 adet marka ismi (K2R, Glory ve Bissell), 3 adet fiyat uygulaması (\$1,19 - \$1,39 - \$1,59) ve Good Housekeeping mührü, para iade garantisi soruları için de “evet/hayır” değişkenlerini kullanmıştır. Bu değişkenlerle 10 tüketici için 22 profil tanımlanmıştır. Tercih değişkeni her profilin içeriğine göre, 1’den 22’ye doğru değer almıştır. En küçük olan sıra en yüksek tercihe aittir. Bu değişkenle her profilin tercihleri yansıtılmaktadır.

Tablo 4. 2: Çalışmada Kullanılan Değişkenler

| Değişken Adı | Değeri |
|--------------------------------|----------------------------|
| Tüketici satın alma tercihleri | 1-22 |
| Ambalaj dizaynı | A, B, C |
| Marka ismi | K2R, Glory, Bissell |
| Fiyat | (\$1,19 - \$1,39 - \$1,59) |
| Good Housekeeping mührü | Evet/Hayır |
| Para iade garantisi | Evet/Hayır |

3. 2. Çalışmanın Yöntemi

Hiyerarşik kluster (küme) analizi adı verilen istatistik yöntemini kullanarak benzer özelliğe sahip objeleri veya olayları gruplandırarak incelemek mümkündür. Analize konu olan her olay birbirine benzeme ya da benzememe eğiliminde olduğundan, kümeleme analizi yöntemiyle birbirine benzer homojen grupların tespiti matematiksel olarak yapılabilmektedir. Diğer istatistik tekniklerinde olduğu gibi bu yöntemde de analizler sırasında hangi veri tiplerinin kullanılacağı, mesafenin hangi matematiksel metotla hesap edileceği ve hangi kriterler kullanılarak grup sayısının belirleneceğine önceden karar verilmektedir (Karabulut vd., 2004:68).

Kümeleme analizinde grupların oluşmasında değişkenler arasındaki mesafe dikkate alınmaktadır. Mesafe ise objelerin veya olayların birbirine göre konumlarını ifade ederken, benzerlik ve yakınlık durumunu da gösterir. Birbirine benzer objeler için mesafe ölçüsü küçük, benzerlik ölçüsü ise yüksek olur. En çok kullanılan indeks Kareli Öklid mesafesidir ki bu da bütün değişkenlerin farklarının (mesafe) karelerinin toplamına eşittir. Değişkenlerin ölçek seviyeleri (ölçü birimleri) sonucu doğrudan etkilediği için ham verilerin standardize edilmiş bir forma getirilmesi gerekmektedir (Gürbüz ve Karabulut, 2009:34). Bu nedenle araştırmada değerler “0” ile “1” arasında standardize edilmiştir. Pearson yakınlık matrisi kullanılarak da profiller arasındaki yakınlıklar tespit edilmiştir. Yakınlık matrisinde mesafe +1 ile -1 arasında değişmektedir. Tüketici satın alma tercihleri bakımından profiller birbirlerine yaklaştıkça yakınlık matrisi +1’e uzaklaştıkça ise -1’e yaklaşmaktadır. Bu çalışmada mesafe ölçüsü olarak Kareli Öklid mesafesi ve Pearson Yakınlık Matrisi kullanılmıştır.

Tüketici satın alma tercihleri arasındaki benzerlik düzeylerini belirlemek amacıyla standardize edilen değerler Aglomeratif Hiyerarşik yöntem kullanılarak gruplandırılmıştır.

Küme veya grupların kombinasyonu sırasında ise Basit Bağlantı ve En Yakın Komşu Tekniği tercih edilmiştir. Ayrıca Pearson Yakınlık Matrisi kullanılarak da profiller arasındaki yakınlıklar belirlenmiştir.

Hiyerarşik kluster sonuçlarını göstermede dendrogram yöntemi de kullanılmıştır. Dendrogram soldan sağa 0–25 birim olarak ölçeklendirilmiştir. Birimler arasındaki mesafeler eşittir. Dendrogramdaki yatay çizgiler mesafeyi, dikey çizgiler ise birleşen kümeleri göstermektedir. Ölçek üzerinde kümelerin birleşme noktaları, hangi grupların oluştuğunu gösterdiği gibi aynı zamanda aralarındaki mesafeyi de göstermektedir. Araştırmada bu yöntemin uygulanması sonucunda tüketici satın alma tercihleri 0–25 birim arasındaki ölçeklendirilmiş mesafelerde gruplandırılmıştır.

10 tüketici için oluşturulan 22 profil arasında birbirlerine en çok benzeyen profilleri ortaya çıkarmak amacıyla uygulanan Hiyerarşik Kümeleme Analizi sonuçlarına göre en anlamlı kümeleme beşli sınıflandırma sonucunda elde edilmiştir. Sonuçlar Tablo 4.2’de verilmiştir.

Aglomeratif çizelgede ise katsayı (coefficients) kullanılarak birbirlerine en çok benzeyen profiller eşleşmiştir. Bu profiller Tablo 4.3’te gösterilmektedir. Bu metodun sonuçlarına göre birbirlerine en çok benzeyen (0.252) profiller, 1. sırada Profil 6 ve Profil 22 olurken 2. sırada Profil 4 ve Profil 12, 3. sırada ise Profil 1 ve Profil 14 olmuştur. Birbirine en az benzeyen profiller ise Profil 1 ve Profil 13 olmuştur. Katsayılardan da görüleceği üzere birbirlerine en çok benzeyen profiller ilk basamaklarda bir araya gelirlerken, birbirlerine daha az benzeyen profiller ise daha sonraki basamaklarda bir araya gelmektedirler.

Tablo 4. 2: Hiyerarşik Küme Analizine Göre Oluşan Gruplar

| Kümeler ve Küme Üyeliği | | | | | |
|-------------------------|-----------|-----------|-----------|-----------|-----------|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | Profil 1 | Profil 2 | Profil 3 | Profil 13 | Profil 15 |
| 2 | Profil 7 | Profil 4 | Profil 8 | | |
| 3 | Profil 14 | Profil 5 | Profil 10 | | |
| 4 | Profil 19 | Profil 6 | Profil 17 | | |
| 5 | | Profil 9 | | | |
| 6 | | Profil 11 | | | |
| 7 | | Profil 12 | | | |
| 8 | | Profil 16 | | | |
| 9 | | Profil 18 | | | |
| 10 | | Profil 20 | | | |
| 11 | | Profil 21 | | | |
| 12 | | Profil 22 | | | |

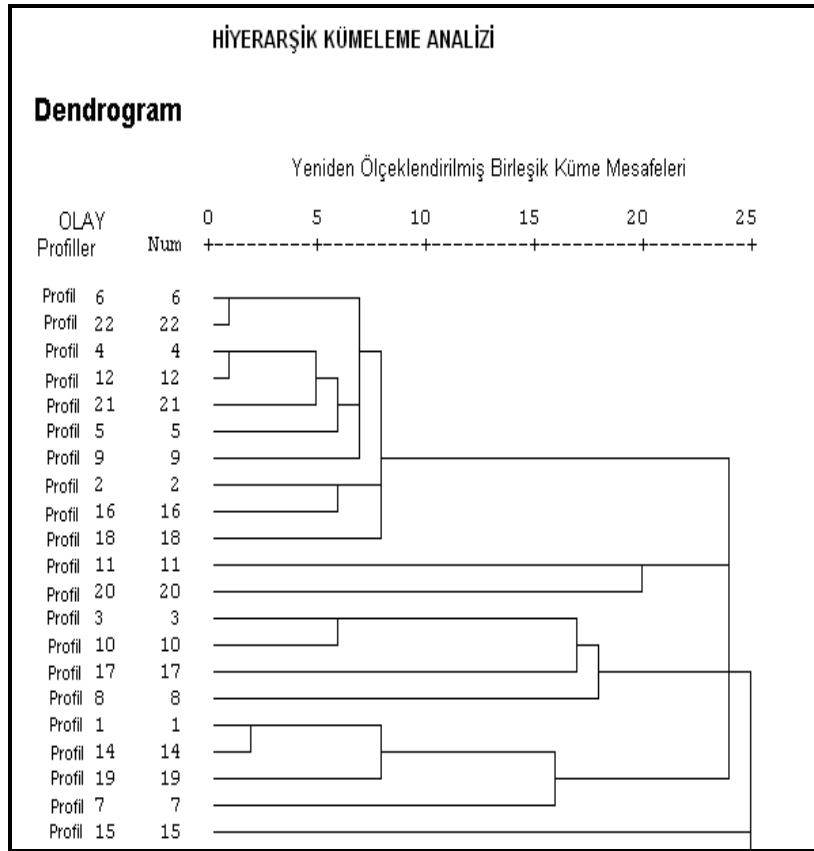
Tablo 4. 3: Aglomeratif Çizelge

(Küme Değerleri Profil Değerlerini Göstermektedir)

| Sıra | Küme 1 | Küme 2 | Coefficients (Katsayı) | Sıra | Küme 1 | Küme 2 | Coefficients (Katsayı) |
|------|----------|-----------|------------------------|------|-----------|-----------|------------------------|
| 1 | Profil 6 | Profil 22 | 0,252 | 12 | Profil 2 | Profil 18 | 0,582 |
| 2 | Profil 4 | Profil 12 | 0,252 | 13 | Profil 1 | Profil 7 | 0,905 |
| 3 | Profil 1 | Profil 14 | 0,307 | 14 | Profil 3 | Profil 17 | 0,977 |
| 4 | Profil 4 | Profil 21 | 0,434 | 15 | Profil 3 | Profil 8 | 1,020 |
| 5 | Profil 2 | Profil 16 | 0,477 | 16 | Profil 11 | Profil 20 | 1,082 |
| 6 | Profil 3 | Profil 10 | 0,502 | 17 | Profil 2 | Profil 11 | 1,252 |
| 7 | Profil 4 | Profil 5 | 0,509 | 18 | Profil 2 | Profil 3 | 1,252 |
| 8 | Profil 4 | Profil 9 | 0,536 | 19 | Profil 1 | Profil 2 | 1,252 |
| 9 | Profil 4 | Profil 6 | 0,536 | 20 | Profil 1 | Profil 15 | 1,332 |
| 10 | Profil 1 | Profil 19 | 0,557 | 21 | Profil 1 | Profil 13 | 1,332 |
| 11 | Profil 2 | Profil 4 | 0,557 | | | | |

Hiyerarşik kümeleme sonuçlarını göstermede kullanılan dendrogram incelendiğinde kullanılan değişkenler bakımından birbirlerine en çok benzeyen profiller 1 birimlik mesafede grup oluştururken, birbirlerine en az benzeyen profiller ise en son aşama olan 25 birimlik mesafede bir araya gelmiştir (Şekil 4.1). Bazı profiller birbirlerine çok benzer olduklarından 1 birimlik mesafede grup oluştururken, bazı profiller kendilerine has özelliklerinden dolayı ilk aşamada tek başlarına bir grup gibi görünmektedirler.

Şekil 4.1: Dendrogram



Profil 6 - Profil 22 ve Profil 4 - Profil 12 arasında benzerlikler çok kuvvetli olduğu için bu profiller 1 birimlik mesafede grup oluşturmaktadırlar. Profil 1 ve profil 14 de birbirlerine çok benzeyen profiller durumundadır, ancak bunlar dendrogramda 2 birimlik mesafede bir araya gelmektedirler. Profil 21, 5 birimlik mesafede Profil 4 ve Profil 12'nin bulunduğu gruba dahil olmuştur. Profil 2 ve Profil 16, 6 birimlik mesafede birbirlerine benzerlik gösterirken; Profil 5, bu mesafede Profil 4, 12 ve 21'in grubuna dahil olmuştur. 7 birimlik mesafede, Profil 6 ve Profil 9 benzerlik gösterirken, Profil 4'ün olduğu grupla birleşmiştir. Profil 18, 8 birimlik mesafede Profil 4'ün ve Profil 2'nin grubuna katılırken, Profil 19 da Profil 1'in grubuna katılmıştır. 16 birimlik mesafede Profil 7 Profil 1'in grubuna dahil olmakta ve 1. kümeyi oluştururken; Profil 17, 17 birimlik mesafede, Profil 8 de 18 birimlik mesafede Profil 3'ün grubuyla birleşmiş ve 3. kümeyi oluşturmuştur. Profil 20, 20 birimlik mesafede Profil 11 ile benzerlik göstermektedir. 25 birimlik mesafe en az benzerlik olan mesafeyi göstermekte ve Profil 20'nin grubuna dahil olarak 1. kümeyi oluşturmaktadır. Profil 15 ve 13, sırasıyla 4. ve 5. kümeyi oluştururken bunlar 25 birimlik mesafeyi de aşmaktadırlar.

Bu bulgulara göre en çok benzerlik gösteren Profil 6 ve Profil 22'nin özellikleri şöyledir. Her iki tüketici satın alma tercihinde de marka Bissell olarak belirtilmiş, Good Housekeeping mührü ve Para iadesinin olması sorularına “hayır” yanıtı vermişlerdir. Profil 4 ve Profil 12’de ise her iki tercih grubu da C amblemini, Bissel markasını seçmişler ve Para iadesine “hayır” yanıtı vermişlerdir. En az benzerlik gösteren Profil 1ve Profil 13 sadece marka tercihinin B olması konusunda ortak özellik göstermektedir.

SONUÇ

Verinin bilgiye dönüşümünün geleneksel yöntemi, klasik çözümleme ve yorumlamaya dayanmaktadır. Ancak günümüzde veri miktarındaki olağanüstü artış, sağlık, eğitim, ticaret, askeri alanlar, alışveriş, devlet sektörü, özel sektör ve pek çok alanda verilerin islenmesi ve bu verilerin değerlendirilerek bilgiye dönüştürülmesi bir zorunluluk haline gelmiştir. Böylece bu alanlardaki gelişmeler daha hızlı ve verimli bir şekilde gerçekleşebilmektedir. Veri madenciliği bu noktada uygun teknikler kullanarak gizli kalmış bilgileri ortaya çıkarma özelliğiyle gereklidir.

Veri madenciliği teknikleri diğer birçok alanda olduğu gibi pazarlama alanında da yaygın bir şekilde kullanılmaktadır. Bu çalışmada, veri madenciliği tekniklerinden kümeleme analizi tekniği kullanılarak SPSS Sample Files’ dan elde edilen veriler üzerinde yöntemin uygulanabilirliği gösterilmiştir. Çalışmada yeni bir halı temizleyicisini pazarlayan bir işletmenin verileri üzerinde kümeleme analizi kullanılarak benzerlikler ortaya çıkarılmıştır. Metodun sonuçlarına göre birbirlerine en çok benzeyen (0.252) profiller 1. sırada Profil 6 ve Profil 22 olurken 2. sırada Profil 4 ve Profil 12, 3. sırada ise Profil 1 ve Profil 14 olmuştur. Birbirine en az benzeyen profiller ise Profil 1 ve Profil 13 olmuştur. En çok benzerlik gösteren Profil 6 ve Profil 22 tüketici satın alma tercihinde de marka Bissell olarak belirtilmiş, Good Housekeeping mührü ve Para iadesinin olması sorularına “hayır” yanıtı vermişlerdir. Profil 4 ve Profil 12’de ise her iki tercih grubu da C amblemini, Bissel markasını seçmişler ve Para iadesine “hayır” yanıtı vermişlerdir. En az benzerlik gösteren Profil 1ve Profil 13 sadece marka tercihinin B olması konusunda ortak özellik göstermektedir.

İşletmeler aynı anda toplumdaki her kesime hitap edemeyeceklerinde tüketici gruplarını belirlemek zorundadırlar. Pazar dilimlemesi yaparak hangi kitleye hitap edeceğini seçmelidir. Bu sayede kendisine farklı bir avantaj sağlayabilmektedir. Bu bağlamda

kümeleme analizi yol göstermede uygun bir teknik olmuştur. Benzerlikleri ortaya çıkarması nedeniyle bulgular kümeleme analizi tekniğinin ihtiyaca yönelik cevapları ortaya çıkarmakta etkili olduğunu göstermektedir. Bununla birlikte kümeleme analiziyle ortaya çıkan sonuçlar farklı yöntemlerde veri ve değişken olarak da kullanılabilmekte ve bu yolla yeni bilgiler elde edilebilmektedir.

KAYNAKÇA

- Akın, Y.K.** (2008). *Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi*, Yayınlanmamış Doktora Tezi, İstanbul: Marmara Üniversitesi, Sosyal Bilimler Enstitüsü.
- Albayrak, A.** ve **Yılmaz, Ş.K.** (2009). “Veri Madenciliği: Karar Ağaçları ve İMKB Verileri Üzerine Bir Uygulama”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1), ss. 31–52.
- Arslan, F.M.** (2004). “Spor Ayakkabısı Satın Alma ve Kullanım Amaçlarına İlişkin Pazar Bölümlerinin Oluşturulması: Üniversite Öğrencileri Üzerine Bir Araştırma”, *Marmara Üniversitesi İİBF Dergisi*, 19(1), ss. 251–276.
- Aşan, Z.** (2007). “Kredi Kartı Kullanan Müsterilerin Sosyo-Ekonomik Özelliklerinin Kümeleme Analiziyle İncelenmesi”, *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, Nisan 2007 (17), ss. 256–268.
- Atbaş, A.C.G.**(2008). *Kümeleme Analizinde Küme Sayısının Belirlenmesi Üzerine Bir Çalışma*, Yayınlanmamış Yüksek Lisans Tezi, Ankara: Ankara Üniversitesi, Fen Bilimleri Enstitüsü.
- Canitez, M.** ve **Yeniçeri, T.** (2007). “İhracat Performansı Düşük Olan Kobi’ler ile Yüksek Olan Kobi’lerin Farklılıklarını Belirlemeye Yönelik Bir Pilot Araştırma”, *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Temmuz-Aralık 2007 (29), ss. 113–134.
- Çetin, E.İ.** (2003). “Çok Değişkenli Analizlerin Pazarlama ile İlgili Araştırmalarda Kullanımı: 1995–2002 Arası Yazın Taraması”, *Akdeniz İ.İ.B.F. Dergisi*, 5, ss. 32–47.
- Day, G.S.** and **Heler, R.M.** (1971). “Using Cluster Analysis to Improve Marketing Experiments”, *Journal of Marketing Research*, 8(3), pp. 340–347.
- Demiralay, M.** ve **Çamurcu, A.Y.** (2005). “CURE, AGNES ve K-MEANS Algoritmalarındaki Kümeleme Yeteneklerinin Karşılaştırılması”, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 4(8), ss.1–18.
- Doğan, B.** (2008). *Bankaların Gözetiminde Bir Araç Olarak Kümeleme Analizi: Türk Bankacılık Sektörü İçin Bir Uygulama*, Yayınlanmamış Doktora Tezi, İstanbul: Kadir Has Üniversitesi Sosyal Bilimler Enstitüsü.
- Durmuş, M.S.** ve **İplikçi, S.** (2007). “Veri Kümeleme Algoritmalarının Performansları Üzerine Karşılaştırmalı Bir Çalışma,” *Akademik Bilişim 2007*, 31 Ocak - 02 Şubat, Kütahya: Dumlupınar Üniversitesi.
- Ergün, E.** (2008). *Ürün Kategorileri Arasındaki Satış İlişkinin Birliktelik Kuralları ve Kümeleme Analizi ile Belirlenmesi ve Perakende Sektöründe Bir Uygulama*, Yayınlanmamış Doktora Tezi, Afyonkarahisar: Afyon Kocatepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Funfgeld, B.** ve **Wang, M.** (2009). “Attitudes and Behaviour in Everyday Finance: Evidence from Switzerland”, *International Journal of Bank Marketing*, 27(2) pp. 108–128.
- Green, P., Frank R.** and **Robinson P.** (1967). “Cluster Analysis in Test Market Selection”, *Management Science*, 13(8), pp.398.
- Green, P.E.** and **Wind, Y.** (1973). *Multiattribute Decisions in Marketing: A Measurement Approach*, SPSS 11.0 Help-Topics-Sample Files.
- Gürbüz, M.** ve **Karabulut, M.** (2009). “SSCB’nin Dağılımıyla Bağımsızlığına Kavuşan Ülkelerde Sosyo-Ekonomik Benzerlik Analizi”, *Bilig, Yaz / 2009(50)*, pp.31-50.
- Hand, D., Manilla, H.** and **Symth, P.** (2001). *Principles of Data Mining*, London: The MIT Press.
- Harrigan, K.R.** (1985). “An Application of Clustering for Strategic Group Analysis”, *Strategic Management Journal*, 6(1), pp. 55-73.
- Helgesen, O., Nettet, E.** and **Voldsund T.** (2009). “Marketing Perceptions and Business Performance Implications for Marketing Education?”, *Marketing Intelligence & Planning*, 27(1), pp. 25-47.
- Karabulut, M., Gürbüz, M.** ve **Sandal, E.K.** (2004). “Hiyerarşik Kluster (Küme)Teknikleri Kullanılarak Türkiye’deki İllerin Sosyo-Ekonomik Benzerliklerinin Analizi”, *Coğrafi Bilimler Dergisi*, 2(2), ss.65-78.
- Ketchen, D.Jr.** ve **Shook, C.L.** (1996). “The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique”, *Strategic Management Journal*, 17(6), pp. 441-458.
- Klastorn, T.D.** (1983). “Assessing Cluster Analysis Results”, *Journal of Marketing Research*, 20(1), pp. 92-98.

- Koyuncugil, A.S.** ve **Özgülbaş, N.** (2006). “İMKB’de İşlem Gören KOBİ’lerin Finansal Başarısızlığına Etki Eden Faktörlerin Veri Madenciliği ile Belirlenmesi”, 3. KOBİ ve Verimlilik Kongresi, 17–18 Kasım, İstanbul.
- Mehrizi, M.H.R.** and **Bontis, N.** (2009). “A Cluster Analysis of The Km Field”, *Management Decision*, 47(5), pp. 792-805.
- Özdamar, K.** (2004). *Paket Programlar İle İstatistiksel Veri Analizi 2*, Eskişehir: Kaan Kitabevi.
- Özekes, S.** (2003). “Veri Madenciliği Modelleri ve Uygulama Alanları”, *İstanbul Ticaret Üniversitesi Dergisi*, 2(3), ss. 65-82.
- Sandal, E. K., Karabulut, M.** ve **Gürbüz, M.** (2005). “Sosyo-Ekonomik Kriterler Bakımından Türkiye’nin Konumu ve Avrupa Birliği”, *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 15(1), ss.1-14.
- Sackson, M.** (1990). “The Use of Cluster Analysis for Business Game Performance Developments in Business”, *Simulation & Experiential Exercises*, 17, pp. 150-154.
- Sambamoorthi, N.** (1999). “Hierarchical Cluster Analysis Some Basics and Algorithms”, Erişim Tarihi: 10.03.2010, http://www.crmportals.com/hierarchical_cluster_analysis.pdf.
- SPSS 15.0** Help Menüsü.
- Ünlükaplan, Y.** (2008). *Çok Değişkenli İstatistiksel Yöntemlerin Peyzaj Ekolojisi Araştırmalarında Kullanımı*, Yayınlanmamış Doktora Tezi, Adana: Çukurova Üniversitesi Fen Bilimleri Enstitüsü.
- Vatanserver, M.** (2008). *Görsel Veri Madenciliği Tekniklerinin Kümeleme Analizlerinde Kullanımı ve Uygulanması*, Yayınlanmamış Yüksek Lisans Tezi, İstanbul: Yıldız Teknik Üniversitesi Fen Bilimler Enstitüsü.
- Yılmaz, Ş.K.** (2008). *Veri Madenciliği: İstanbul Menkul Kıymetler Borsası Örneği*, Yayınlanmamış Yüksek Lisans Tezi, Zonguldak: Zonguldak Karaelmas Üniversitesi Sosyal Bilimler Enstitüsü.
- Yücel, İ.** ve **Yeniçeri, H.** (2008). “İhracat Yapan Kobi’lerin İnsan Sermayesi, Örgütsel Sermaye, Dışsal Sermaye İle İhracat Performansı İtibariyle Gruplandırılması”, *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 20, s. 707-721.
- Yüksel, C.A.** (2002). “Dondurulmuş Hazır Yemek Satın Alan Tüketicilerin Özelliklerini Belirlemeye Yönelik Araştırma”, *Ege Akademik Bakış Dergisi*, 2(1), s.1-12.