

VERİ MADENCİLİĞİNDE KÜMELEME YAKLAŞIMLARI VE KOHONEN AĞLARI İLE PERAKENDECİLİK SEKTÖRÜNDE BİR UYGULAMA

CLUSTERING APPROACHES IN DATA MINING AND AN APPLICATION WITH KOHONEN NETWORKS IN RETAILING SECTOR

Arş.Gör.Dr. Çağatan TAŞKIN*
Yrd.Doç.Dr. Gül Gökay EMEL**

ÖZET

Bu çalışmada; bir perakende işletmenin müşterilerinin Kohonen ağları ile kümelenmesi ele alınmaktadır. Kohonen ağlarının seçilmesinin nedeni, büyük hacimli veriler üzerinde çalışabilme yetisi ve kümeleme analizi için en önemli karar olan küme sayısını, tekniğin kendisinin en uygun olarak belirleyebilmesidir. Kümeleme analizinin amacı; ele alınan işletmeye, pazar bölümlendirmesi ve hedef pazar seçimi gibi stratejik pazarlama kararlarında yardımcı olması için, önceden bilinmeyen kritik müşteri özellikleri ve önem derecelerini de ortaya çıkararak gerekli öngörüleri sağlamaktır.

ABSTRACT

In this study; clustering of a retailer's customers is handled by means of Kohonen networks. Kohonen networks is chosen because of its ability to work on large data and to find out the number of clusters which is the most important decision of a cluster analysis. The aim of this cluster analysis is to provide strategic information to the retailer in order to help in its market segmentation and target market selection decisions while discovering previously unknown critical customer attributes and their importance.

Veri Madenciliği, Kümeleme, Kohonen Ağları, Pazar Bölümlendirme
Data Mining, Clustering, Kohonen Networks, Market Segmentation

1. GİRİŞ

Günümüzde; giderek artan ve çeşitlenen tüketici gereksinim ve isteklerine paralel olarak, bilgi ve iletişim teknolojilerindeki hızlı gelişim,

* Uludağ Üniversitesi, İ.İ.B.F., İşletme Bölümü, Bursa.

** Uludağ Üniversitesi, İ.İ.B.F., İşletme Bölümü, Bursa.

pazarlamanın gerçekleştiriliş biçimini, işletmelerin müşterileri ile ilgili veri ve bilgileri yönetme biçimini değiştirmiştir. Bazı bilgi teknolojileri araçları sayesinde toplanabilen müşteriler ile ilgili büyük hacimli veriler, işletmelerin rekabet avantajları elde edebilmesi için önemli fırsatlar içermektedir. Ancak; bu verilerin stratejik kararları desteklemede ve etkin pazarlama stratejilerini oluşturmada kullanılabilmesi için hassas, analitik ve çok boyutlu olarak analiz edilmesi gerekir. Veri madenciliği bu işlemi yapan en önemli araçlardan biridir. Çeşitli veri madenciliği araçları günümüzde pazarlamacılara, uygun pazarlama kararları alabilmeleri için doğru özbilgiyi sağlayabilmektedir.

İşletmelerin stratejik pazarlama kararlarından biri de pazar bölümlendirme ile ilgilidir. Özellikle, işletmelerin mevcut ya da hedef pazarlarının anlamlı ve etkin pazar bölümlerine ayrılabilmesi son derece kritik bir stratejik karardır. Bu çalışmanın amacı; bir perakende işletmenin müşterilerinin bir veri madenciliği tekniği olan Kohonen ağları ile kümelendiğini ele alarak, işletmenin pazar bölümlendirme ve hedef pazar seçimi gibi stratejik pazarlama kararlarını almasında, önceden bilinmeyen kritik müşteri özellikleri ve önem derecelerini ortaya çıkararak işletmeye gerekli öngörüü sağlamaktır. Kohonen ağları, doğal küme sayısını bulmada iyi bir yöntem olması ve büyük hacimli veriler üzerinde etkin sonuçlar üretmesi nedenleri ile analiz tekniği olarak seçilmiştir.

2. VERİ MADENCİLİĞİ VE KÜMELEME

Veritabanlarının sayı ve hacimlerindeki hızlı artış ile birlikte stratejik karar almayı desteklemek ve güvenilir tahmin yapabilmek için veriler içinden anlamlı bilgi elde edilmesi giderek önem kazanmaktadır. Bu işlevi yerine getiren araç veri madenciliğidir¹. Bir başka ifade ile veri madenciliği, büyük miktarlarda depolanmış veriler içinde saklı kuralların, örüntülerin ve modellerin ortaya çıkarılması sürecidir². Veri madenciliği bu süreci bilgisayar bilimini, makine öğrenmesini, veritabanı yönetimini, matematiksel algoritmaları ve istatistiği birleştirerek gerçekleştirir³. Veri yığınları içindeki gizli bilgi, işletmelerin stratejik planlar oluşturmada yararlı ipuçları sağlayabilir^{4,5}. Veri madenciliği, temel veri erişimi, istatistik,

¹ Chih-Ping WEI, Yen-Hsien LEE ve Che-Ming HSU, “Empirical Comparison of Fast Partitioning-Based Clustering Algorithms For Large Data Sets”, **Expert Systems With Applications**, Vol:24, Issue:4, 2003, s. 351.

² Michael J. SHAW, Chandrasekar SUBRAMANIAM, Gek Woo TAN, ve Michael E. WELGE, “Knowledge Management And Data Mining For Marketing”, **Decision Support Systems**, Vol:1, Issue:31, 2001, s. 128.

³ Shu-hsien LIAO., “Knowledge Management Technologies And Applications—Literature Review From 1995 To 2002”, **Expert Systems With Applications**, Vol: 25, No:2, 2003, s. 157.

⁴ Eden W.M. MA ve Tommy W.S. CHOW, “A New Shifting Grid Clustering Algorithm”, **Pattern Recognition**, Vol:37, Issue:3, 2004, s. 503.

⁵ John A. McCARTY ve Manaj HASTAK, “Segmentation Approaches in Data Mining: A Comparison of RFM, CHAID and Logistic Regression”, **Journal of Business Research**, Vol:60, Issue: 6, 2009, s. 656.

OLAP gibi diğer veri analizi tekniklerini tamamlayıcı bir yaklaşımdır. Veri madenciliği ile boyut bazında gruplama, boyutlar arasında ilişkiyi inceleme, veri görselleştirme ve sonuçları grafik ve rapor olarak sunma olanağı sağlanır. Üretilen özbilgi, özellikle stratejik karar vermede işletme yöneticilerinin iyi bir yardımcısıdır⁶.

Veri madenciliğinde tahmin edici modeller ile örüntü tanıma işi sınıflama, regresyon ve zaman serileri yaklaşımlarını içerir. Bu modeller, neyin tahmin edilmesinin istendiğine dayalı olarak farklılaşırlar. Çıktı niteliğinin sürekli değerleri için tahmin istenir ise regresyon analizi, zamanın ayırt edici özellikleri ile ilgileniliyor ise zaman serileri, iyi veya kötü gibi az sayıdaki ayrık kategoriye sahip bir özel veri ögesi için bir kategorik değer tahmini yapılmak isteniyor ise sınıflama gerekir. Eldeki verinin gruplarını bulan kümeleme, birliktelik ve ardışıklık kurallarını elde etmeyi kapsayan birliktelik analizi ve ardışıklık keşfi davranışı ise tanımlama amaçlı kullanılır^{7,8}.

Kümeleme, birincil amacı özelliklerine göre nesnelere gruplamak olan çok değişkenli bir analiz tekniğidir. Kümelemede nesnelere gruplanması önceden belirlenmiş bir kritere göre gerçekleştirilir. Kümeleme sonuçları, kümeler içinde yüksek derecede homojenlik, kümeler arasında ise yüksek derecede heterojenlik göstermelidir. Kümeleme yapısı içerisinde rastlantı değişkeni kavramının önemli bir yeri vardır. Ancak, rastlantı değişkeninin kümeleme yapısı içindeki yeri diğer çok değişkenli tekniklerde olduğundan tamamen farklıdır. Küme rastlantı değişkeni, kümelemede nesnelere karşılaştırmak için kullanılan özellikleri simgeleyen değişkenler kümesidir. Küme rastlantı değişkeni yalnızca nesnelere karşılaştırmak için kullanılan değişkenleri içerdiği için nesnelere özelliğini belirler. Kümeleme, rastlantı değişkenini deneysel olarak tahmin etmeyen ve ayrıca rastlantı değişkeninin araştırmacı tarafından belirlendiği tek çok değişkenli analiz tekniğidir. Kümelemenin odak noktası rastlantı değişkeninin tahmini değil, rastlantı değişkenine dayalı olarak nesnelere karşılaştırılmasıdır^{9,10}.

Genellikle kümeleme denetimsiz bir biçimde gerçekleşir. Denetimsiz öğrenme yetisinden dolayı, veriler içinde gizli olan örüntüleri ortaya çıkarmayı sağlar. Bu durum kümelemeyi, örüntü tanıma,

⁶ Gül Gökay EMEL, Çağatan TAŞKIN ve Serhat KILIÇARSLAN, "Sinir Ağları Veri Madenciliği İle Çelik Üretim Sürecinde Bir Analiz", **Dokuz Eylül Üniversitesi İşletme Fakültesi Dergisi**, Cilt:5, Sayı:1, 2004, s. 206-207.

⁷ Chris RYGIELSKI, Jyun-Cheng WANG ve David C. YEN, "Data mining Techniques For Customer Relationship Management", **Technology in Society**, Vol:24, Issue: 4, 2002, s. 488.

⁸ Helen M. MOSHKOVICH, Alexander I. MECHITOV ve David L. OLSON, "Rule Induction In Data Mining: Effect Of Ordinal Scales", **Expert Systems With Applications**, Vol:22, Issue:4, 2002, s. 303-304.

⁹ Cenk BUDAYAN, İrem DİKMEN ve M. Talat BİRGÖNÜL, "Comparing the Performance of Traditional Cluster Analysis, Self-organizing Maps and Fuzzy C-means Method for Strategic Grouping", **Expert Systems with Applications**, Vol:36, Issue:9, 2009, s. 11772.

¹⁰ F. Joseph HAIR, E. Rolph ANDERSON, L. Ronald TATHAM ve C. William BLACK, **Multivariate Data Analysis With Readings**, 4th Edition, Prentice-Hall, USA, 1995, s. 423.

biyoinformatik ve veri madenciliği gibi çeşitli alanların önemli bir araştırma konusu haline getirmiştir¹¹.

3. KÜMELEME ALGORİTMALARI

Literatürde birçok kümeleme algoritması mevcuttur. Kümeleme algoritmaları; hiyerarşik ve hiyerarşik olmayan olmak üzere iki genel sınıfa ayrılarak ele alınmaktadır. Bazı kümeleme algoritmaları ise birden fazla sınıfa ait olabilir. Hiyerarşik sınıfın iki alt sınıfı mevcuttur. Bunlar; yığılmacı (agglomerative) ve bölücü (divisive) yaklaşımlarıdır. Hiyerarşik olmayan sınıf ise ayırma (partitioning), yoğunluk-tabanlı (density-based), ızgara-tabanlı (grid-based) ve diğer yaklaşımlar olmak üzere dört alt sınıfa ayrılmaktadır. Eğer kullanılan kümeleme algoritması aşamalı olarak kümeler oluşturuyor ve veriyi kümelerden oluşan bir ağaç şeklinde gösteriyor ise hiyerarşik sınıfa aittir. Yığılımsal sınıf alttan-üste olarak adlandırılan bir yaklaşımdır. Bu yaklaşımda, veri noktaları birbiri ile birleştirilerek kümeler oluşturulur. Bölücü yaklaşım ise tersine üstten-alta olarak adlandırılan bir işleyişe sahiptir. BIRCH, CURE, CHAMELEON ve ROCK algoritmaları hiyerarşik sınıfa ait algoritmalarlardır. Hiyerarşik sınıfa ait algoritmalar, birbirine en fazla benzeyen iki nesneyi bir kümede toplar. Bu, işlem maliyeti oldukça yüksek bir süreçtir. Çünkü her toplamadan önce tüm nesnelere karşılaştırılır. Bölücü yaklaşımın da gerektirdiği hesaplama yükü yığılımsal yaklaşıma benzer. Ayrıca, hiyerarşik kümeleme yöntemlerine örnek olarak Ward'ın minimum varyans yöntemi de verilebilir¹².

Hiyerarşik olmayan sınıfa giren kümeleme algoritmaları veriyi doğrudan kümeleyen algoritmalarlardır. Ayırma tipi kümeleme, hiyerarşik olmayan kümeleme sınıfında en yaygın kullanılan yaklaşımdır. Bu tip algoritmalar genellikle, tüm noktalar ilgili küme merkezleri ile uzaklıkları minimuma ininceye kadar kümelerin merkezini değiştirir. Ayırma yaklaşımının en yaygın örneği olarak “K-means” verilir. “Bulanık C-means”, “K-modes” ve “K-prototypes” gibi algoritmalar “K-means” algoritmasının çeşitleridir¹³. Hiyerarşik sınıf ile karşılaştırıldığında ayırma yaklaşımı hesaplama zamanı bakımından daha düşük maliyetlidir. Ancak, tüm noktaların merkezleri ile mesafesi minimum olana kadar, her nokta ile ilgili merkezi arasındaki mesafe ardışılmsal olarak hesaplandığı için gene de yüksek maliyetli bir süreçtir. Ayırma yaklaşımının uzaklık ölçütü öklidyen uzaklık gibi uzaklık formülüne dayanır. Bu durum, bu tip algoritmaların küresel şekilli kümeleri bulmadaki yetisini sınırlar. Ayrıca, bu tip algoritmalar rastsal şekilli kümelerin bulunmasında da sorun yaşarlar.

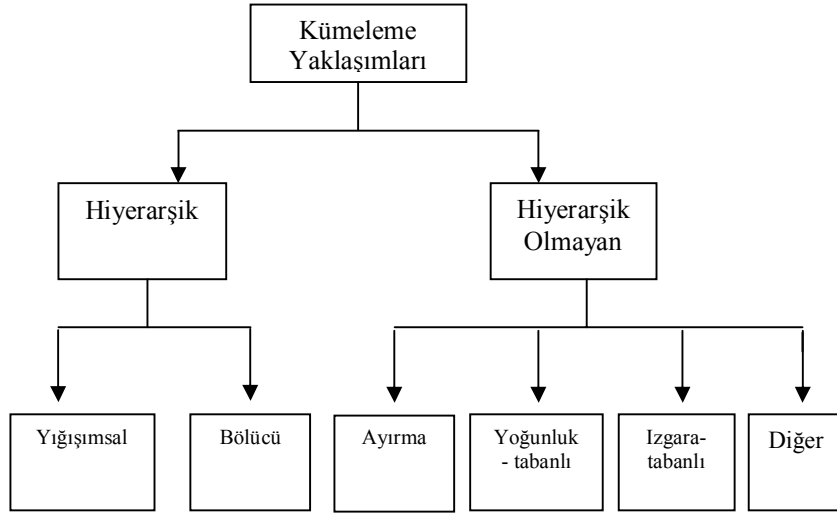
¹¹ MA ve CHOW, s. 503.

¹² MA ve CHOW, s. 503-504; R.J. KUO, L.M. HO ve C.M. HU, “Cluster Analysis In Industrial Market Segmentation Through Artificial Neural Network”, **Computers&Industrial Engineering**, Vol:42, Issue:2-4, 2002, s. 393.

¹³ Aristidis LIKAS, Nikos VLASSIS ve Jakob J. VERBEEK, “The Global K-Means Clustering Algorithm”, **Pattern Recognition**, Vol:36, Issue:2, 2003, s. 451.

Yoğunluk-tabanlı kümeleme yaklaşımı ise diğer bir hiyerarşik olmayan kümeleme yaklaşımıdır. Yoğunluk-tabanlı kümeleme algoritması verinin yoğun olduğu alanı küme olarak aldığı için, rastsal şekillere sahip kümeler bulmada sorun yaşamaz. Tipik algoritma örnekleri arasında; DBSCAN, DENCLUE ve OPTICS algoritmaları sayılabilir. Izgara-tabanlı kümeleme yaklaşımı veri noktalarından çok hücreleri göz önüne alan bir yaklaşımdır. Bu özelliğinden dolayı, ızgara-tabanlı kümeleme algoritmaları genel olarak tüm kümeleme algoritmalarından hesapsal olarak daha etkindir. Bu yaklaşıma örnek olarak; STING, STING+, WaveCluster, CLIQUE ve GDILC verilebilir. Yukarıda belirtilen tipteki algoritma sınıfları içine girmeyen algoritmalar ise “diğer” sınıfı içinde yer alırlar¹⁴.

Şekil 1: Kümeleme Yaklaşımlarının Sınıflandırılması



Kaynak: Eden W.M. MA ve Tommy W.S. CHOW, “A New Shifting Grid Clustering Algorithm”, *Pattern Recognition*, Vol:37, Issue:3, 2004, s. 504.

3.1. İki Aşamalı Kümeleme Algoritmaları

İki aşamalı kümeleme algoritmaları denildiğinde, ilk akla gelen Punj ve Steward (1983) tarafından önerilen klasik iki aşamalı kümeleme algoritmasıdır. Bu algoritma, Ward’ın minimum varyans yöntemi ile “K-means” yönteminden oluşan bir hibrid yaklaşımdır. Böyle bir karma yaklaşımın avantajı, Ward’ın minimum varyans yönteminin, “K-means” yönteminin gerektirdiği küme sayısını hesaplamasından ileri gelmektedir^{15,16}.

¹⁴ Chih-Hung HSU, “Data Mining to Improve Industrial Standards and Enhance Production and Marketing: An Empirical Study in Apparel Industry”, *Expert Systems with Applications*, Vol:36, Issue:3, 2009, s. 4186; MA ve CHOW, s. 504.

¹⁵ R.J. KUO, L.M. HO ve C.M. HU, “Integration of Self-Organizing Feature Map and K-means Algorithm For Market Segmentation”, *Computers&Operations Research*, Vol: 29, Issue: 11, 2002, s. 1476.

Ayrıca, Kuo ve arkadaşlarının yapay sinir ağları ile çok değişkenli analizi bütünleştiren bir diğer iki aşamalı kümeleme algoritması mevcuttur. Bu kümeleme yaklaşımında, kendini düzenleyen haritalar (self-organizing maps) ve K-means teknikleri iki aşamalı bir yaklaşım içerisinde kullanılmaktadır. Kendini düzenleyen haritalar tekniği ilk aşamada küme sayısının bulunmasında, K-means algoritması ise son çözüme ulaşılmasında kullanılmaktadır. Araştırma sonuçlarına göre, önerilen iki aşamalı bu yöntemin klasik iki aşamalı yöntemden (Ward'ın minimum varyans yöntemi ve K-means yöntemi) daha iyi sonuçlar verdiği belirtilmektedir¹⁷.

Klasik iki aşamalı kümeleme yaklaşımlarında, birinci aşamada hiyerarşik tekniklerin kullanıldığı görülmektedir. Ancak bu tekniklerin geri alınamama gibi bir dezavantajı mevcuttur. Başka bir ifade ile bir gözlem bir kümeye atanması durumunda artık başka bir kümeye atanamaz. Ancak, Kohonen yaklaşımı öğrenen bir algoritma niteliğindedir ve sürekli olarak güncellenebilir. Böylece, Kohonen yaklaşımı ile uygun küme sayısı elde edilebilir¹⁸.

3.2. Kohonen Ağları

Kohonen tarafından önerilen “kendini düzenleyen haritalar” geniş bir uygulama alanına sahip olan veri madenciliği tekniklerinden biridir¹⁹. Kendini düzenleyen haritalar, yapay sinir ağları tekniklerinden biri olan standart ileri besleme ağlarından sonra en yaygın kullanılan teknik olup, kümeleme ve boyut indirgeme için kullanılan denetimsiz iki katmanlı bir yapay sinir ağı olarak tanımlanabilir²⁰. Bilindiği üzere yapay sinir ağları, insan beyninden esinlenerek geliştirilmiş, ağırlıklı bağlantılar aracılığıyla birbirine bağlanan ve her biri kendi belleğine sahip işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır²¹.

Kohonen ağları giriş ve çıkış nöronlarından oluşur. Giriş nöronlarının sayısını veri setindeki değişken sayısı belirler. Çıkış nöronlarının her biri bir kümeyi temsil eder. Diğer yapay sinir ağlarından farklı olarak, çıkış katmanındaki nöronların dizilimi çok önemlidir. Bu dizilim doğrusal, dikdörtgen, altıgen veya küp şeklinde olabilir. En çok

¹⁶ Moreno M.N. GARCIA, Luis A. Miguel QUINTALES, Francisco J. Garcia PENALYO ve M. Jose Polo MARTIN, “Building Knowledge Discovery-Driven Models for Decision Support in Project Management”, **Decision Support Systems**, Vol: 38, Issue: 2, 2004, s. 309.

¹⁷ KUO vd., s. 394-395.

¹⁸ KUO vd., s. 1481.

¹⁹ Barbara HAMMER, Alessio MICHELI, Alessandro SPERDUTI ve Marc STRICKERT, “Recursive Self-Organizing Network Models”, **Neural Networks**, Vol:17, Issue:10, 2004, s. 1061.

²⁰ Dmitri G. ROUSSINOV ve Hsinchun CHEN, “Document Clustering For Electronic Meetings: An Experimental Comparison Of Two Techniques”, **Decision Support Systems**, Vol: 27, Issue: 1-2, 1999, s. 70.

²¹ Murat ÇUHADAR, İbrahim GÜNGÖR ve Ali GÖKSU, “Turizm Talebinin Yapay Sinir Ağları İle Tahmini ve Zaman Serisi Yöntemleri İle Karşılaştırmalı Analizi: Antalya İline Yönelik Bir Uygulama”, **Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, Cilt:14, Sayı:1, 2009, s. 102.

dikdörtgenel ve altıgen şeklindeki dizilimler tercih edilmektedir. Uygulamada, çoğu kez dikdörtgenel dizilim karesel dizilim olarak uygulanır. Buradaki dizilim topolojik komşuluk açısından önemlidir. Aslında, çıkış nöronları arasında doğrudan bir bağlantı yoktur. Giriş nöronları ile her bir çıkış nöronu arasındaki bağlantıyı referans vektörleri (code-book vectors) gösterir. Bu vektörler bir katsayılar matrisinin sütunları olarak da düşünülebilir. Kohonen ağları eğitilirken, bu topolojik komşuluk referans vektörlerinin yenilenmesinde kullanılır. Bu ağlarda kullanılan öğrenme algoritması denetimsizdir. Yani, ağ eğitilirken bağımlı değişken kullanılmaz²². Bir başka ifade ile, denetimsiz öğrenmede girdi değişkenlerine karşılık arzu edilen çıktılar belirtilmez. Ağ yalnızca girdi modelini öğrenir. Öğrenme süreci üzerindeki ileri dönüşün kaynağı belli değildir. Katmanlar arasındaki ağırlıkların ayarlanması ağ tarafından kendiliğinden gerçekleştirilir. Denetimsiz öğrenmede sadece girdi katmanındaki değerler kullanılmaktadır. Amaç, veri setindeki modelin ortaya çıkarılmasıdır. Sistemin doğru çıkış hakkında bilgisi yoktur ve girişlere göre kendi kendisini örnekler. Denetimsiz olarak eğitilebilen ağlar, istenen ya da hedef çıkış olmadan giriş bilgilerinin özelliklerine göre ağırlık değerlerini ayarlar. Burada ağ istenen dış verilerle değil, girilen bilgilerle çalışır. Bu tür öğrenmede gizli sinirler dışarıdan yardım almaksızın kendilerini örgütlemek için bir yol bulmalıdırlar. Bu yaklaşımda, verilen giriş vektörleri için önceden bilinebilen performansını ölçebilecek ağ için hiçbir çıkış örneği sağlanmaz, yani ağ kendisi uygulayarak öğrenmektedir²³.

Kendini düzenleyen haritalar, standart kümeleme algoritmalarından farklı bir yapıya sahiptirler. Temel farklılık, bu yöntemin ayrı kümeler üzerinde işlem yapması yerine, birbiri ile ilişkili veri noktalarını gruplara ataması şeklinde ifade edilebilir²⁴. Bu tekniğin diğer kümeleme algoritmalarına göre bir avantajı, çok boyutlu verileri iki boyutlu bir ızgara kullanarak görselleştirmesidir. Çok boyutlu ölçekleme tekniğine benzer bir teknik olduğu söylenebilir. Kendini düzenleyen haritalar tekniğinde, her girdi düğümü bir boyuta karşılık gelirken, her çıktı düğümü de iki boyutlu ızgarada bir düğümüne karşılık gelir²⁵.

Tablo 1’de bazı kümeleme yaklaşımlarının bir karşılaştırması verilmektedir.

²² Hüdaverdi BİRCAN, Metin ZONTUL ve A.Gürkan YÜKSEK, “SOM Tipinde Yapay Sinir Ağlarını Kullanarak Türkiye’nin İhracat Yaptığı Ülkelerin Kümelenmesi Üzerine Bir Çalışma”, *Atatürk Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt:20,Sayı:2, 2006, s.222-223.

²³ Metehan TOLON, *Tüketici Tatmininin Yapay Sinir Ağları Yöntemiyle Ölçülmesi ve Ankara’daki Perakendeci Mağazaların Müşterileri Üzerinde Bir Uygulama*, Gazi Üniversitesi Sosyal Bilimler Enstitüsü, Yayınlanmamış Doktora Tezi, Ankara, 2007, s.77.

²⁴ B. CURRY ve Peter Huw MORGAN, “Evaluating Kohonen’s Learning Rule: An Approach Through Genetic Algorithms”, *European Journal of Operational Research*, Vol:154, Issue: 1, 2004, s. 191-192.

²⁵ Melody Y. KIANG, Dorothy M. FISHER, Jeng-Chung Victor CHEN, Steven A. FISHER ve Robert T. CHI, “The Application of SOM as a Decision Support Tool to Identify AACSB Peer Schools”, *Decision Support Systems*, Vol:47, Issue:1, 2009, s. 51-52; ROUSSINOV ve CHEN, s. 70.

Tablo 1: Bazı Kümeleme Yaklaşımlarının Karşılaştırılması

Kümeleme Yaklaşımları	Avantajlar	Dezavantajlar
Yapay Sinir Ağları (Kohonen)	Doğal başlangıç Büyük hacimli veriler üzerinde çalışabilme Küme sayısını belirleyebilmesi	Uzun hesaplama zamanları Eğitim parametrelerinin ayarlanmasının zor olması ve farklı parametrelerin farklı sonuçlar vermesi
Hiyerarşik Yöntemler (Ward'ın Minimum Varyans Yöntemi)	Küme sayısını belirleyebilmesi	Büyük hacimli veriler üzerinde çalışma zorluğu Aykırı değerlerden (outlier) kolayca etkilenmesi Geri alınamama
Hiyerarşik Olmayan Yöntemler (K-means Yöntemi)	Başlangıç noktası ve küme sayısı verildiğinde daha yüksek doğruluk Büyük hacimli veriler üzerinde çalışabilme	Küme sayısını belirleyememe Başlangıç noktasının ve küme sayısının rastsal olarak seçilmesi Aynı gruba ait iki merkezin seçilebilmesi

Kaynak: R.J. KUO, L.M. HO ve C.M. HU, "Integration of Self-Organizing Feature Map and K-means Algorithm For Market Segmentation", **Computers&Operations Research**, Vol: 29, Issue: 11, 2002, s. 1478.

4. PERAKENDE İŞLETMESİ MÜŞTERİLERİNİN KOHONEN AĞLARI İLE KÜMELENMESİ

Bu kısımda, gerçekleştirilen analizin amacı, kapsamı, yöntemi, adımları ve sonuçları yer almaktadır.

4.1. Analizin Amacı

Bu çalışmada; sebze, meyve, taze et, süt ürünleri, konserve sebze, konserve et, konserve yemek, balık, bira, şarap, alkolsüz içki ve gıyecek satışı yapan bir perakende işletmenin belirli bir dönem boyunca alışveriş yapmış müşterilerinin Kohonen tekniği ile kümelenebileceği ele alınmaktadır. Kohonen tekniğinin kümeleme analizi aracı olarak seçilmesinin nedeni, büyük hacimli veriler üzerinde çalışabilme yetisi ve kümeleme analizi için en önemli karar olan küme sayısını, tekniğin kendisinin en uygun olarak belirleyebilmesidir. Kümeleme analizin amacı; ele alınan işletmenin veritabanında yer alan müşterilerin veritabanında yer alan tüm özelliklere göre kümelere ayrılmasını sağlayarak, işletmenin pazar bölümlendirmesi ve hedef pazar seçimi gibi stratejik pazarlama kararlarını almasında işletmeye gerekli öngörüü sağlamaktır. Ayrıca, yapılan kümeleme analizi ile

işletmenin mevcut pazarının anlamlı ve etkin pazar bölümlerine ayrılabilmesi için kritik müşteri özellikleri ve önem dereceleri de ortaya konmaktadır.

4.2. Veri Seti ve Analiz Yöntemi

Bu çalışmada, yapılan analiz için Kohonen ağları tekniği kullanılmakta ve Clementine v8.1 veri madenciliği yazılımından yararlanılmaktadır. Clementine programı veri akışına dayanmaktadır. Bu akış içinde düğümler ile gösterilen her bir işlem, birbirine oklar ile bağlanmaktadır. Analizler, 6 aylık bir dönemde satın alınan ürünleri ile tutarlarını ve müşterilere ait bazı özellikleri içeren işletmenin 10000 adet müşteri kaydından oluşan bir müşteri veri tabanı üzerinde gerçekleştirilmektedir. Müşterilerin çoğunluğu bu dönemde birden fazla alışveriş gerçekleştirmiştir. Bununla beraber, aynı müşteriye ait her bir alışveriş kaydı birbirinden farklı olabilmektedir. Örneğin, X müşteri numarasına sahip tüketici, bir alışverişinde “konserve sebze” alırken, başka bir alışverişinde almayabilir. Dolayısıyla da, aynı müşteri numarasına sahip kayıtlardan birinde “konserve sebze”, “evet” olarak kodlanırken, diğerinde “hayır” olarak kodlanmıştır. Öz bir ifade ile her bir müşteri alışverişi veri tabanında ayrı bir kayıt olarak gözükmektedir. Veri tabanında işletmenin satış verileri müşteriye göre düzenlenmiştir. Müşteri kayıtlarında yer alan değişkenler Tablo 2’de verilmektedir:

Tablo 2: Müşteri Kayıtlarındaki Değişkenler

DEĞİŞKENLER	
Müşteri Numarası	Sebze/Meyve
Aylık Satış Tutarı	Taze Et
Ödeme Tipi	Süt Ürünleri
Cinsiyet	Konserve Sebze
Evsahibi	Konserve Et
Aylık Gelir	Konserve Yemek
Medeni Durum	Hane Halkı Sayısı
Yaş	Bira
Alkolsüz İçki	Şarap
Balık	Giyecek

4.3. Analizin Adımları

Analize verilerin programa yerleştirilmesi ile başlanmıştır. Ele alınan işletmeye ait veri tabanı Clementine yazılımındaki “Source” düğümü yardımı ile okutulur. Değişkenlerin tipini belirlemek için ise “Type” düğümü kullanılır. Ek 1’de, değişken tiplerine ve değişkenlerin girdi olarak seçimine bir örnek verilmektedir. Burada, Kohonen analizi için tüm değişkenler girdi olarak seçilmiştir. Böylece tüm özelliklere göre doğal kümeler elde edilmesi amaçlanmaktadır. Kümeleme yapılabilmesi için, “Kohonen” düğümü seçilerek, “Type” düğümüne eklenir. “Kohonen” düğümü ise Ek 2’de

verilmektedir. “Kohonen” düğümünde yer alan “execute” düğmesi tıklandıktan sonra verilerin eğitimi süreci başlamaktadır.

4.4. Bulgular

“Kohonen” modeli sonucunda sekiz küme elde edilmiştir. Bu kümeler ve kümeleme analizi sonucunda önemli bulunan nitelikler Tablo 3’te verilmektedir. Elde edilen kümelerden beş tanesinin ayrıntılı özellikleri ise Tablo 4’te verilmektedir. Tablo 4’te yer alan “küme 1”, “küme 2”, “küme 3” ve “küme 4” en büyük dört küme olması, “küme 5” ise diğer kümelere göre daha farklı bir alışveriş kalıbının bulunması nedeni ile seçilmiştir.

Tablo 3: Değişkenlerin Önem Dereceleri

DEĞİŞKENLER	
Sebze/Meyve: Önemsiz	Taze Et: Önemsiz
Aylık Satış Tutarı: Önemsiz	Süt Ürünleri: Önemsiz
Ödeme Tipi: Önemsiz	Konserve Sebze: Önemli
Cinsiyet: Marjinal	Konserve Et: Önemli
Ev sahibi: Önemli	Konserve Yemek: Önemli
Aylık Gelir: Önemli	Hane Halkı Sayısı: Önemsiz
Medeni Durum: Önemsiz	Bira: Önemli
Yaş: Önemsiz	Şarap: Önemsiz
Alkolsüz İçki: Önemsiz	Giyecek: Önemsiz
Balık: Önemli	

Tablo 3’te, kümeleme için kullanılan girdilerden “ev sahibi olma”, “aylık gelir”, “balık”, “konserve sebze”, “konserve et”, “konserve yemek” ve “bira” değişkenlerinin kümeleme için önemli olduğu, cinsiyet değişkeninin ise önemli olmakla beraber, diğer girdilere göre daha az önemli olduğu (programda marjinal olarak adlandırılmaktadır) görülebilir. Belirtilen değişkenler dışında kalanlar ise önemsiz olarak bulunmuştur. Aşağıda yer alan Tablo 4’te kümeleme analizi sonucunda bulunan en önemli nitelikler ve önem dereceleri verilmektedir.

Tablo 4: Kohonen Analizi Sonuçları

En Önemli Nitelikler (Önem Dereceleri)	Küme 1 (X = 3, Y = 2)	Küme 2 (X = 3, Y = 0)	Küme 3 (X = 0, Y = 0)	Küme 4 (X = 0, Y = 1)	Küme 5 (X = 0, Y = 2)
Aylık Gelir (1.0)	2000-3000 TL	1000-2000 TL	3000 TL üzeri	2000-3000 TL	1000 TL'den az
Ev Sahibi Olma (1.0)	Evet (%70)	Evet (%60)	Evet (%60)	Hayır (%55)	Evet (%60)
Balık (1.0)	Evet (%80)	Evet (%70)	Evet (%60)	Evet (%60)	Evet (%60)
Konserve Sebze (1.0)	Evet (%90)	Evet (%60)	Evet (%100)	Evet (%60)	Hayır (%70)
Konserve Yemek (1.0)	Evet (%80)	Evet (%60)	Evet (%90)	Evet (%60)	Hayır (%70)
Bira (1.0)	Evet (%90)	Evet (%60)	Evet (%90)	Evet (%60)	Hayır (%70)
Konserve Et (0.99)	Evet (%70)	Evet (%70)	Evet (%80)	Evet (%80)	Evet (%70)
Cinsiyet (0.95)	Erkek (%55) Kadın (%45)	Erkek (%70) Kadın (%30)	Erkek (%60) Kadın (%40)	Erkek (%45) Kadın (%55)	Erkek (%50) Kadın (%50)
Küme Büyüklüğü	(%20)	(%18)	(%16)	(%13)	(%11)

Tablo 4'teki Kohonen analizi sonuçlarına göre, söz konusu işletme için pazar bölümlendirmede kullanılacak en önemli değişkenlerden birinin tüketicilerin "aylık geliri" olduğu görülebilir. Birinci küme, "orta gelirli" bir pazar bölümü olarak nitelenebilir. Bu pazar bölümündeki tüketicilerin %70'inin kendi evi vardır. "Balık", "konserve sebze" ve "bira" tüketim oranları oldukça yüksektir. İkinci küme ise, "orta-alt gelirli" olarak nitelenebilir. %70'i erkek tüketicilerden oluşmaktadır. Üçüncü küme "yüksek gelirli" ve %60'ı kendi evi olan tüketicilerden oluşan bir pazar bölümüdür. Üçüncü kümenin %60'ı erkeklerden oluşmaktadır. Bu pazar bölümünün diğer pazar bölümlerine göre daha fazla "konserve ağırlıklı" bir beslenme biçimine sahip olduğu söylenebilir. Ayrıca bira ve balık da satın alma örüntüleri içinde yer almaktadır. Dördüncü kümeye bakıldığında "orta gelirli" ve %55'i kadın tüketicilerden oluştuğu görülmektedir. Ayrıca, üçüncü küme gibi "konserve yiyecek" satın alma örüntüsüne rastlanmıştır. Beşinci küme ise "düşük gelirli" ve genel olarak "konserve ağırlıklı" beslenme biçimi olmayan tüketicilerden oluşmaktadır. Diğer iki kümeden önemli farkı alkollü içecek satın alma örüntüsüne diğer kümelere göre çok daha az olmasıdır. Kohonen ağları analizi ile söz edilenler gibi başka bilgiler de elde edilebilir.

5. SONUÇ

Günümüzün ağır rekabet koşulları altında; rekabetçi avantaj elde edebilmek için, işletmelerin potansiyel tüketicilerine ya da müşterilerine ait verilerin düzenli bir şekilde, hassas, doğru ve çok boyutlu olarak analiz edilmesi gerekir. Özellikle pazar bölümlendirme ve hedef pazar seçimi gibi stratejik kararlar çok boyutlu analiz yapabilen modern karar destek araçları yardımı ile alınmalıdır. Bu araçlardan biri olan veri madenciliği çeşitli teknikleri sayesinde gizli bazı örüntüleri ortaya çıkararak, işletmelere rekabetçi üstünlük sağlamada çok önemli bazı bilgiler sağlayabilir.

Bu çalışmada, bir perakende işletmenin belirli bir dönem boyunca alışveriş yapmış müşterilerinin Kohonen tekniği ile kümelenebileceği ele alınmıştır. Yapılan kümeleme analizi ile işletmenin mevcut pazarının anlamlı ve etkin pazar bölümlerine ayrılabilmesi için önceden bilinmeyen kritik müşteri özellikleri ve önem dereceleri de ortaya çıkarılmıştır. Elde edilen bu ve benzeri sonuçlar, işletmenin pazar bölümlendirmesi ve hedef pazar seçimi gibi stratejik pazarlama kararlarını almasında yardımcı olarak kullanılabilir. Çalışmada, kümeleme analizi için Kohonen tekniğinin tercih edilmesinin en önemli iki nedeni ise, uygun küme sayısını doğal olarak hesaplayabilme özelliği ve büyük hacimli veriler üzerinde çalışabilme yetisidir. Bununla birlikte, Kohonen tekniğinin küme sayısını girdi olarak kullanan teknikler ile bileşik olarak kullanılması durumunda daha etkin kümeleme sonuçları elde edilebilir. Benzer konularda yapılması düşünülen çalışmalarda, özellikle iki aşamalı tekniklerin uygulanmasına ağırlık verilebilir.

KAYNAKÇA

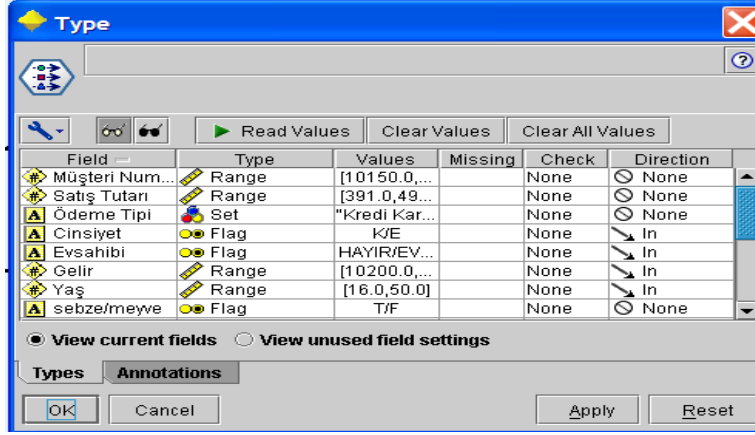
1. BİRCAN, Hüda-verdi, ZONTUL, Metin ve YÜKSEK, A.Gürkan, “SOM Tipinde Yapay Sinir Ağlarını Kullanarak Türkiye'nin İhracat Yaptığı Ülkelerin Kümelene-mesi Üzerine Bir Çalışma”, **Atatürk Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, Cilt:20,Sayı:2, 2006, s. 219-237.
2. BUDAYAN, Cenk, DİKMEN, İrem ve BİRGÖNÜL, M.Talat, “Comparing the Performance of Traditional Cluster Analysis, Self-organizing Maps and Fuzzy C-means Method for Strategic Grouping”, **Expert Systems with Applications**, Vol:36, Issue:9, 2009, s. 11772-11781.
3. CURRY, B. ve MORGAN, Peter Huw., “Evaluating Kohonen's Learning Rule: An Approach Through Genetic Algorithms”, **European Journal of Operational Research**, Vol:154, Issue: 1, 2004, s. 191–205.
4. ÇUHADAR, Murat, GÜNGÖR, İbrahim ve GÖKSU, Ali, “Turizm Talebinin Yapay Sinir Ağları İle Tahmini ve Zaman Serisi Yöntemleri İle Karşılaştırmalı Analizi: Antalya İline Yönelik Bir Uygulama”,

Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt:14, Sayı:1, 2009, s. 99-114.

5. EMEL, Gül Gökay, TAŞKIN, Çağatan ve KILIÇARSLAN, Serhat, “Sinir Ağları Veri Madenciliği İle Çelik Üretim Sürecinde Bir Analiz”, **Dokuz Eylül Üniversitesi İşletme Fakültesi Dergisi**, Cilt:5, Sayı:1, 2004, s. 205-225.
6. GARCIA, Moreno, M.N., QUINTALES, Luis A. Miguel, PENALYO, Francisco J. Garcia ve MARTIN, M. Jose Polo, “Building Knowledge Discovery-Driven Models For Decision Support In Project Management”, **Decision Support Systems**, Vol: 38, Issue: 2, 2004, s. 305-317.
7. HAIR, F. Joseph, ANDERSON, E. Rolph, TATHAM, L. Ronald ve BLACK, C. William, **Multivariate Data Analysis With Readings**, 4th Edition, Prentice-Hall, USA, 1995.
8. HAMMER, Barbara, MICHELI, Alessio, SPERDUTI, Alessandro ve STRICKERT, Marc, “Recursive Self-Organizing Network Models”, **Neural Networks**, Vol:17, Issue:10, 2004, s. 1061–1085.
9. HSU, Chih-Hung, “Data Mining to Improve Industrial Standards and Enhance Production and Marketing: An Empirical Study in Apparel Industry”, **Expert Systems with Applications**, Vol:36, Issue:3, 2009, s. 4185-4191.
10. KIANG, Melody Y., FISHER, Dorothy M., CHEN, Jeng-Chung Victor, FISHER, Steven A. ve CHI, Robert T., “The Application of SOM as a Decision Support Tool to Identify AACSB Peer Schools”, **Decision Support Systems**, Vol:47, Issue:1, 2009, s. 51-59.
11. KUO, R.J., HO, L.M. ve HU, C.M., “Integration of Self-Organizing Feature Map and K-means Algorithm For Market Segmentation”, **Computers&Operations Research**, Vol: 29, Issue: 11, 2002, s. 1475-1493.
12. KUO, R.J., HO, L.M. ve HU, C.M., “Cluster Analysis In Industrial Market Segmentation Through Artificial Neural Network”, **Computers&Industrial Engineering**, Vol:42, Issue:2-4, 2002, s. 391-399.
13. LIAO, Shu-hsien, “Knowledge Management Technologies And Applications—Literature Review From 1995 To 2002”, **Expert Systems With Applications**, Vol: 25, No:2, 2003, s. 155–164.

14. LIKAS, Aristidis, VLASSIS, Nikos ve VERBEEK, Jakob, J., “The Global K-Means Clustering Algorithm”, **Pattern Recognition**, Vol:36, Issue:2, 2003, s. 451-461.
15. MA, Eden, W. M. ve CHOW, Tommy, W.S., “A New Shifting Grid Clustering Algorithm”, **Pattern Recognition**, Vol:37, Issue:3, 2004, s. 503-514.
16. McCARTY, John A. ve HASTAK Manaj, “Segmentation Approaches in Data Mining: A Comparison of RFM, CHAID and Logistic Regression”, **Journal of Business Research**, Vol:60, Issue: 6, 2009, s. 656-662.
17. MOSHKOVICH, Helen M., MECHITOV, Alexander I. ve OLSON, David L., “Rule Induction In Data Mining: Effect Of Ordinal Scales”, **Expert Systems With Applications**, Vol:22, Issue:4, 2002, s. 303-311.
18. ROUSSINOV, Dmitri G. ve CHEN, Hsinchun, “Document Clustering For Electronic Meetings: An Experimental Comparison Of Two Techniques”, **Decision Support Systems**, Vol: 27, Issue: 1-2, 1999, s. 67-79.
19. RYGIELSKI, Chris, WANG, Jyun-Cheng ve YEN, David C., “Data mining Techniques For Customer Relationship Management”, **Technology in Society**, Vol:24, Issue: 4, 2002, s. 483-502.
20. SHAW, Michael J., SUBRAMANIAM, Chandrasekar, TAN, Gek Woo ve WELGE, Michael E., “Knowledge Management And Data Mining For Marketing”, **Decision Support Systems**, Vol:1, Issue:31, 2001, s. 127-137.
21. TOLON, Metehan, **Tüketici Tatmininin Yapay Sinir Ağları Yöntemiyle Ölçülmesi ve Ankara’daki Perakendeci Mağazaların Müşterileri Üzerinde Bir Uygulama**”, Gazi Üniversitesi Sosyal Bilimler Enstitüsü, Yayınlanmamış Doktora Tezi, Ankara, 2007.
22. WEI, Chih-Ping, LEE, Yen-Hsien ve HSU, Che-Ming, “Empirical Comparison of Fast Partitioning-Based Clustering Algorithms For Large Data Sets”, **Expert Systems With Applications**, Vol:24, Issue:4, 2003, s. 351-363.

Ek 1: Kohonen Analizi İçin Girdi Seçimine Bir Örnek



Ek 2: "Kohonen" Düğümü

