

# KÜMELEME ANALİZİNDE YENİ BİR YAKLAŞIM: KENDİNİ DÜZENLEYEN HARİTALAR (KOHONEN AĞLARI)

Ayşe OĞUZLAR<sup>(\*)</sup>

**Özet:** Kendini düzenleyen haritalar (Self-Organizing Maps-SOM) veya diğer bir söyleyişle Kohonen ağı 1980' lerde geliştirilen en önemli ağ yapılarından birisidir. SOM' un ana fonksiyonu, n-boyutlu uzaydaki girdi verilerini, orijinal topolojik ilişkilerini koruyarak daha düşük boyutlu (genellikle bir veya iki) çizimle haritalaştırmasıdır. Bu çalışmada Bursa Emniyet Müdürlüğünden sağlanan veriler ile SOM, suçlu profilinin belirlenmesi amacıyla kullanılmıştır. Suçlu profilleri tanımlamaları C5.0 kural algoritması değerlendirilerek yapılmıştır.

**Anahtar Kelimeler:** Kendini düzenleyen haritalar, Kohonen ağıları.

**Abstract:** Self-organizing maps (SOM) or Kohonen's network is one of the most important network architectures developed during the 1980s. The main function of SOM is to map the input data from an n-dimensional space to a lower dimensional (usually one or two) plot while maintaining the original topological relations. In this study with cooperation of police headquarters of Bursa, SOM is used for definition of profiles of criminals. As a result criminal profiles were defined according to evaluation of C5.0 rule algorithm.

**Keywords:** Self-organizing maps, Kohonen's network.

## I.Giriş

Veri madenciliğinin üstlendiği görevler geniş bir bakış açısıyla dört boyutta incelenebilir: Sınıflama, tahmin, bölümlendirme/kümeleme, tanımlama/özetleme. Görevin ana hedefine bağlı olarak, veri madenciliği sürecinin çıktısı ön kestirimsel modeller veya betimleyici bilgiler olacaktır. Veri madenciliği süreci tarafından türetilen ön kestirimsel modeller sınıflama veya tahmin görevleri söz konusu olduğunda kullanışlı iken; betimleyici bilgiler, bölümlendirme/kümeleme ve özetleme türünden problemler için uygun olacaktır. Her bir veri madenciliği uygulaması farklı hedef ve durumlara sahip olduğundan, farklı veri madenciliği tekniği kümelerini gerektirmektedir.Tablo 1' de veri madenciliğinin amaçları, görevleri ve teknikleri/araçları özetlenmiştir.

---

<sup>(\*)</sup> Yrd.Doç.Dr. Uludağ Üniversitesi İİBF Ekonometri Bölümü

Tablo 1: *Veri Madenciliği Amaçları, Görevleri ve Teknikleri/Araçları*  
(Kiang ve Kumar: 2001, 180)

Amaçlar	Görevler	Teknikler/Araçlar
Ön Kestirim	Sınıflama	Logit modeller, diskriminant analizi, k-en yakın komşuluk, karar ağaçları, sinir ağları ve genetik algoritma.
	Tahmin	Regresyon, logit modeller, sinir ağları ve genetik algoritma.
Betimleme	Bölümlendirme/Kümeleme	K-ortalama analizi, sinir ağları ve genetik algoritma
	Özetleme/Link Analizi	Birliktelik kuralları, sorgulama araçları (SQL), ANOVA, basit çapraz tablolar, daire grafiği ve histogram gibi görselleştirme teknikleri ve çok boyutlu ölçekleme analizi.

Veri madenciliği, veri tabanında beklenmeyen ilişkilerin bulunmasına yardım etmesi açısından geleneksel istatistiksel analizlerden farklılık göstermektedir. Yüksek boyutluluk ve büyük hacimli verilerden ötürü, veri madenciliğinde geleneksel istatistiksel yöntemlerin kullanımı sınırlanmıştır (Wang ve Wang, 2002:31). Dolayısıyla, veri madenciliğinde yapay zeka ve veri görselleştirme alanındaki çalışmaların zorunlu hale geldiğini söylemek yanlış olmayacaktır.

Bu çalışmada, yapay zeka dolayısıyla sinir ağları alanında yer alan, bir veri görselleştirme aracı ve birimlerin kümelendirilmesi amacına hizmet eden SOM veya Kohonen ağları üzerinde durularak, Uludağ Üniversitesi Bilimsel Araştırma Projeleri kapsamı içerisinde yer alan ve Bursa Emniyet Müdürlüğünden sağlanan ağır suçlar bürosuna ilişkin veriler üzerinde bir uygulama yapılmaya çalışılmıştır.

## II. Kendini Düzenleyen Haritalar (Kohonen Ağları)

Kohonen ağları, başlangıçta grupların bilinmediği durumlarda, verilerin kümelendirilmesini amaçlamaktadır. Kümeleme amacıyla kullanılmasının yanında, bir “veri görselleştirme aracı”dır (Flexer, 2001:382). Kohonen ağlarında, kestirim yapılacak bir çıktı (bağlı) değişken bulunmadığından, denetimsiz öğrenim gerçekleştiren bir sinir ağı türüdür (Oğuzlar, 2004:62). Kohonen ağları, girdi (bağımsız) değişken kümesindeki örüntülerin açığa çıkarılması amacıyla kullanılmaktadır. Bu ağın çıktısında, gözlemler gruplandırılmış olarak elde edilmektedir. Bir grup veya kümenin içindeki gözlemlerin birbirine benzer olduğu, farklı gruplarda yer alan gözlemlerin ise birbirine benzer olmadığını söylemek mümkündür. Kohonen ağları, bir girdi ve iki boyutlu bir Kohonen tabakasından oluşmaktadır (Kiang: 2001, 163). Her bir sinir (nöron), girdi

değişkenleri veya aynı anlamda girdi alanlarının her biri ile bağlantılıdır ve tekrar ağırlıkları (önemleri) bu bağlantıların her birinin üzerinde yer almaktadır. Bir sinir için ağırlıklar, analizde kullanılan girdi alanlarının oluşturduğu küme için bir profili temsil etmektedir. Genellikle bir Kohonen ağı, az sayıda birim çok sayıda gözlemi özetlediğinde (güçlü birimler) veya çok sayıda birim gözlemlerin her hangi birine karşılık gelemediğinde (zayıf birimler) son bulacaktır. Kohonen veya kendini düzenleyen harita, bir çıktı olarak düşünülen sınırları içermesine karşılık, bu ağlarda gerçek çıktı katmanı bulunmamaktadır (Clementine 7.0 User' s Guide, 2002:274). Bir Kohonen haritası Şekil 1' de gösterilmiştir.

Kohonen ağında yer alan daha düşük katmanlardaki düğümler (girdi düğümleri), örneklem veri noktaları tarafından temsil edilen girdileri alırlar. Daha yüksek katmanlardaki düğümler (çıkı düğümleri), denetimsiz öğrenim sürecinin ardından girdi örüntülerinin organizasyon haritasını temsil edecektir.

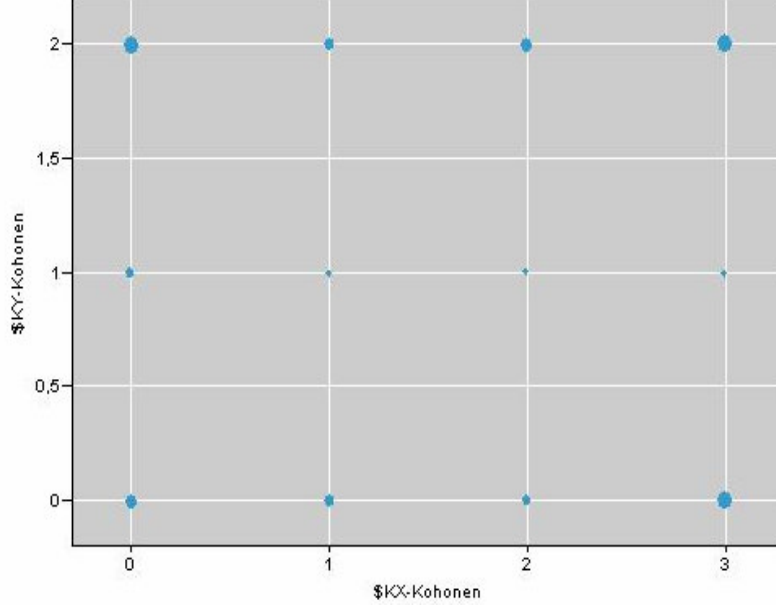
SOM' daki denetimsiz öğrenim süreci şu şekilde tanımlanabilir: Başlangıçta, bağlantı ağırlıkları (connection weights) küçük rassal sayılar ile düzenlenir. Bir örneklem veri noktası tarafından temsil edilen yeni girdi vektörü, girdi düğümleri tarafından alınır. Girdi vektörü, bağlantılara bağlı olarak çıktı düğümlerine gönderilir. Çıkı düğümlerinin faal duruma getirilmesi girdiye bağlıdır. "Kazanan-tümünü-alır (winner-takes-all)" yarışı içerisinde, girdi vektörüne en fazla benzerlik gösteren ağırlıklara sahip çıktı düğümü aktif hale gelecektir. Öğrenim aşamasında ağırlıklar, aşağıdaki Kohonen öğrenim kuralı çerçevesinde güncelenirler:

$$w (\text{yeni})= w (\text{eski})+\eta [X-w(\text{eski})] \quad (1)$$

Bu eşitlikte  $w$  ağırlık matrisi,  $X$  girdi vektörü ve  $\eta$  zaman süresince azalan öğrenim oranı ( $0<\eta<1$ ) dir. Ağırlık güncellemesi yalnızca, aktif çıktı düğümü ve topolojik komşuları için oluşur. Komşuluk geniş olarak başlar ve zaman süresince yavaş bir biçimde azalır.

Girdi vektörlerinin yeterli bir sayısının sunulmasının ardından ağırlıklar, kümeleri belirlemeye başlayacaktır. Kümelerin belirlenmesi ise, küme merkezlerinin lokal yoğunluk fonksiyonlarının, girdi vektörlerinin yaklaşık olasılık yoğunluk fonksiyonlarına yaklaşma eğilimi ile gerçekleşir. SOM' daki çıktı düğümleri, öncül küme merkezleri bilgisi olmadan gerçek kümeleri temsil edecektir (Wang, 2001:129).

Ağın topolojisi, kafes formu olarak tanımlanır. En çok kullanılan iki kafes, dikdörtgen ve altıgen kafeslerdir. Dikdörtgen bir kafeste bir düğüm dört komşuya sahip iken, altıgen bir kafeste ise bu sayı altıdır. Dolayısıyla, görselleştirme amaçları için altıgen kafes tercih edilmektedir (Kloptchenko vd., 2004:31).

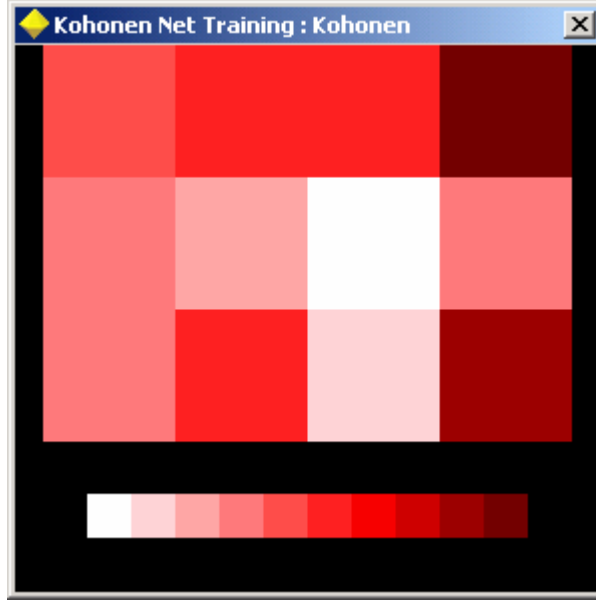


Şekil 1: Kohonen Haritası

Şekil 1 incelendiğinde, analize dahil edilen girdi değişkenlerinden 12 farklı kümenin oluştuğu gözlemlenebilmektedir. Bu 12 farklı kümenin içerdiği gözlem sayılarının birbirinden farklı olduğu, noktaların büyüklüklerinin farklı oluşlarından anlaşılabilir.

Kohonen ağları, boyut azaltma amacına bağlı olarak da kullanılabilir.  $k$  adet orijinal girdi değişkeninden, grafiksel düzenleme sonucunda bulunan iki türetilmiş değişken (şekil 1'de \$KX-Kohonen ve \$KY-Kohonen), orijinal girdi değişkenlerinin benzerlik ilişkilerini korumaktadırlar. Bu özellik, faktör veya temel bileşenler analizinin sağladığı yarara eş değer olmaktadır (Kiang ve Kumar, 2001:178).

Kohonen ağlarının eğitimi boyunca, istenildiği takdirde bir geri bildirim grafiği (feedback graph) görüntülenebilir. Bir Kohonen Geri Bildirim grafiği örneği Şekil 2' de gösterilmiştir.



Şekil 2: Kohonen Geri Bildirim Grafiği

Kohonen geri bildirim grafiği, eğitim süresince görüntülenmektedir. Her bir düğümün gücü, kırmızıdan beyaza doğru değişen bir renk ile temsil edilmektedir. Kırmızı renk, çok sayıda gözlem kazanan birimleri (güçlü birimler), beyaz renk ise birkaç veya hiç kayıt kazanamayan birimleri (zayıf birimler) temsil etmektedirler.

Kümeleme analizinde kullanılan istatistiksel tekniklere alternatif olarak SOM geliştirilmiştir. Kümelemede “benzer” veri noktaları aynı küme içerisinde yer almaktadır (Dasu ve Johson, 2003:77). SOM yöntemi, kümeleme analizinde kullanılan tipik bir yapay zeka tekniğidir. Sinir ağlarına dayalı tekniklerin, istatistiksel modelleme tekniklerine tercih edilmesinin en önemli nedenlerinden biri, verilerin dağılımlarına ilişkin varsayımlar gerektirmemeleridir. SOM yüksek boyutlu verilerin daha düşük boyutlu resmedildiği bir harita sunar ve bu harita da veri madenciliği araştırmacılarına kümeleri görüntüler. SOM yöntemi bazı istatistiksel kümeleme yöntemleri gibi kümelerin ön sayısına, değişkenlerin olasılık dağılımlarına ve değişkenler arasındaki bağımsızlığa ilişkin varsayımlar gerektirmez (Dunham, 2003:148).

SOM, Tuevo Kohonen tarafından tanımlanan bir sinir ağı türüdür. Diğer sinir ağları türlerinde olduğu gibi, veri kümelerindeki ilişkilerin açıklanmasında kullanılan parametrik olmayan bir yaklaşımdır (Penn, 2005; 2). Kohonen sinir ağlarına dayalı SOM, kümeleme analizinde arzu edilir bir teknik haline gelmiştir. SOM’ a dayalı kümeleme teknikleri veri madenciliğinde istatistiksel yöntemlerden daha çok avantajlara sahiptir. Bu avantajlardan ilki, veri

madenciliğinin çok yüksek dereceden veriler ile ilgili olmasından kaynaklanmaktadır. Bu yüksek boyutluluk, verilerin istatistiksel korelasyonlarını anlamsız hale getirmekte ve bu sebeple istatistiksel yöntemler güçsüz kalmaktadırlar. SOM yöntemi, istatistiksel testlerin dayalı olduğu bir takım varsayımlara dayanmamakta ve yüksek boyutlu veriler ile çalışıldığında etkin bir yöntem olarak düşünülmektedir. İkinci olarak, veri madenciliğinde yer alan veriler, bilinen çok değişkenli dağılımlara genellikle sahip değillerdir ve bu nedenle geleneksel istatistiksel kümeleme yöntemleri bu durumlarda sınırlamalara sahiptirler. Diğer taraftan istatistiksel varsayımlara karşı gevşekliği, kullanışlı ve esnek oluşu nedeni ile kümeleme analizi çalışmalarında SOM yöntemi sıkça kullanılmaya başlamıştır. Üçüncüsü, SOM yöntemi yüksek boyutlu veri kümelerinin görselleştirilmesi için bir temel teşkil etmektedir. Bu özellik diğer veri analizi yöntemlerinin hiçbirinde bulunmamaktadır. SOM yönteminin en büyük dezavantajı ise, kümeleme analizi sonuçlarının geçerliliğine ilişkin bir ölçü sağlayamamasıdır. Ayrıca SOM, kümelerin birbirinden ayırt edilmelerini sağlayacak özelliklerin (bağımsız değişkenlerin) neler olduğunu veremediğinden dolayı, bir kural türetme tekniği ile (C5.0 gibi) birlikte kullanılması daha anlamlı olacaktır (Clemetine 7.0 User's Guide, 2002:384). Çalışmamızda da SOM' un elde edilmesinin ardından, oluşan kümelerin özelliklerinin belirlenmesi amacıyla C5.0 kural türetme algoritması kullanılmıştır.

### III. Uygulama

Uygulama bölümünde Bursa Emniyet müdürlüğünden alınan polis suç veri tabanına ilişkin ağır suçlar bürosuna ait aşağıdaki değişkenler ele alınmıştır:

Bağımlı Değişken:

BURO:

Birey ağır suç işlemiş ise 1, diğer suçlardan yakalanmış ise 0 (Bu değişken C5.0 kural türetme algoritmasında kullanılmıştır).

Bağımsız Değişkenler:

Suçun İşlendiği Saat (OLAYSAAT):

1. 00.00-05.59
2. 06.00-11.59
3. 12.00-17.59
4. 18.00-23.59

Suçu İşleyen Bireyin Cinsiyeti (CINSIYET):

1. Erkek
2. Kadın

Suç İŞleyen Bireyin Yaşı (YAS)

Suç İŞleyen Bireyin Doğum Yeri (DOGYER):

1. Akdeniz Bölgesi
2. Doğu Anadolu Bölgesi
3. Ege Bölgesi
4. Güney Doğu Anadolu Bölgesi
5. İç Anadolu Bölgesi
6. Karadeniz Bölgesi
7. Marmara Bölgesi
8. Göçmen (Yurt Dışı)

Suç İŞleyen Bireyin Öğrenim Durumu (OGRENIM):

1. Okur Yazar Değil
2. Okur Yazar
3. İlkokul Mezunu
4. Ortaokul mezunu
5. Lise Mezunu
6. Meslek Yüksekokulu Mezunu
7. Üniversite Mezunu

Suç İŞleyen Bireyin Mesleği (MESLEK):

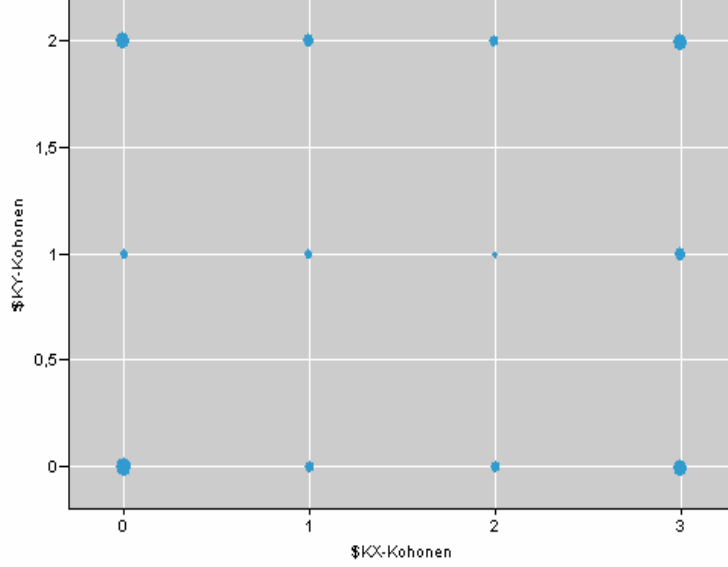
1. Boşta Gezer
2. Öğrenci
3. Çiftçi
4. Ev Hanımı
5. İşçi (Kamu veya Özel)
6. Kamu Personeli (Devlet Memuru)
7. Serbest Meslek (Avukat, Mühendis, Doktor, Muhasebeci, Teknisyen,.....)
8. Esnaf ve Sanatkar
9. Yönetici
10. Emekli

117931 bireyden oluşan veri tabanına öncelikle Celementine 7.0 paket programında yer alan Kohonen modülü uygulandığında aşağıdaki Tablo 2 elde edilmiştir.

Tablo 2: Kohonen Çıktısı

\$KX-Kohonen: 4□
\$KY-Kohonen: 3□
Girdi Katmanı: 35 nöron□
Çıktı Katmanı: 12 nöron□

Tablo 2 incelendiğinde Kohonen analiz modülü uygulanması sonucunda toplam 12 ayrı kümenin oluştuğu, girdi katmanının 35 ve çıktı katmanının da 12 nörona sahip olduğu görülmektedir. Bu 12 kümenin grafiği ise aşağıda Şekil 3'de gösterilmiştir.



Şekil 3: Suçlu Ver Tabanına İlişkin Kohonen Grafiği

Kohonen grafiğinin elde edilmesinin ardından oluşan 12 farklı kümenin özellikleri araştırılmak istenmiştir. Bu amaçla öncelikle C5.0 algoritması kullanılarak, ağır suç işleyen bireylerin X ve Y Kohonen değerleri bulunmaya çalışılmıştır. Ağır suç işleyen bireylerin X ve Y Kohonen değerlerine ilişkin elde edilen iki kural aşağıdaki şekildedir.

Ağır Suç İşleyen Bireyler İçin:

Kural 1: Eğer  $\$KY\text{-Kohonen} \leq 0$  ve  $\$KX\text{-Kohonen} > 0$  ise birey ağır suç işlemektedir

Kural 2: Eğer  $0 < \$KX\text{-Kohonen} \leq 1$  ve  $\$KY\text{-Kohonen} > 0$  ise birey ağır suç işlemektedir.

Diğer Suçları İşleyen Bireyler İçin:



Kural 1: Eğer  $\$KX\text{-Kohonen} > 1$  ve  $\$KY\text{-Kohonen} > 0$  ise birey diğer suçları işlemektedir.

Kural 2:  $\$KX\text{-Kohonen} \leq 0$  ise birey diğer suçları işlemektedir.

Elde edilen kurallar ve Şekil 3.' de gösterilen Kohonen grafiği incelendiğinde; ağır suç işleyen ve birinci kuralı sağlayan bireylerin oluşturduğu kümelerin; ( $\$KX\text{-Kohonen}=1$ ,  $\$KY\text{-Kohonen}=0$ ), ( $\$KX\text{-Kohonen}=2$ ,  $\$KY\text{-Kohonen}=0$ ) ve ( $\$KX\text{-Kohonen}=3$ ,  $\$KY\text{-Kohonen}=0$ ) değerlerine sahip olduğu, ağır suç işleyen ve ikinci kuralı sağlayan bireylerin ise ( $\$KX\text{-Kohonen}=1$ ,  $\$KY\text{-Kohonen}=1$ ) ve ( $\$KX\text{-Kohonen}=1$ ,  $\$KY\text{-Kohonen}=2$ ) değerlerini gerçekleştirdiği anlaşılabilmektedir. Dolayısıyla 12 farklı kümeden oluşan Kohonen grafiğinde sözü edilen 5 farklı ikiliyi gerçekleştiren kümelerin ağır suç işleyen bireylere ilişkin olduğu anlaşılabilmektedir. Ağır suç işleyen bireylere ilişkin kümelerin bulunmasının ardından bu kümelere ilişkin bireylerin ele alınan bağımsız değişkenlere bağlı olarak özelliklerinin araştırılmasına geçilmiştir.

Ağır suç işleyen ve  $\$KX\text{-Kohonen}=1$ ,  $\$KY\text{-Kohonen}=0$  kümesinde yer alan bireylerin ele alınan bağımsız değişkenler açısından özelliklerinin bulunabilmesi için öncelikle  $\$KX\text{-Kohonen}=1$ ,  $\$KY\text{-Kohonen}=0$  değerlerini sağlayan bireyler veri tabanından ayıklanmıştır. Daha sonra ayıklanan bireylere ilişkin olan veri tabanına C5.0 algoritması uygulanarak, bu kümede yer alan ve ağır suç işleyen bireylerin özelliklerine ilişkin kurallar üretilmiştir. Kuralların hemen yanındaki parantezler içerisindeki değerler sırasıyla örnek sayıları (instance-kuralı sağlayan birey sayısı) ve güven değerleri (confidence-İlgili kuralda yer alan ön koşulu sağlayan bireylerin ağır suç işleme yüzdesi) gösterilmiştir. İlk kümeye ilişkin olarak bulunan kurallar şu şekildedir:

**KURAL 1.1** (97, 0,66): Esnaf ve Sanatkar olan ve olayı (00.00-05.59), (06.00-11.59) veya (18.00-23.59) arasında işleyenler ağır suç işlemektedir.

**KURAL 1.2** (404, 0,604): Kadın, Marmara Bölgesi doğumlu, yaşı 25' den büyük olan ve olayı (00.00-05.59), (06.00-11.59) veya (18.00-23.59) arasında işleyenler ağır suç işlemektedir.

**KURAL 1.3** (7.081, 0,542): Olayı (00.00-05.59), (06.00-11.59) veya (18.00-23.59) arasında işleyenler ağır suç işlemektedir.

Ağır suç işleyen ve ( $\$KX\text{-Kohonen}=2$ ,  $\$KY\text{-Kohonen}=0$ ) kümesinde yer alan bireyler için C5.0 algoritması yardımıyla türetilen kurallar ise şu şekilde verilebilir:

KURAL 2.1 (7, 0,857): Kamu Personeli (Devlet Memuru) olan, doğum yeri bilinmeyen veya Akdeniz Bölgesi olan, yaşı 34 veya daha küçük olanlar ağır suç işlemektedir.

KURAL 2.2 (6, 0,833): Doğum yeri bilinmeyen veya Akdeniz Bölgesi olan, serbest meslek sahibi, yaşı 19' dan büyük fakat 31 veya daha küçük olanlar ağır suç işlemektedir.

KURAL 2.3 (495, 0,745): İşçi, erkek ve yaşı 19' dan daha büyük olanlar ağır suç işlemektedir.

KURAL 2.4 (13, 0,769): Doğum yeri bilinmeyen veya Akdeniz Bölgesi olan, esnaf ve sanatkar ve yaşı 19' dan büyük fakat 30 veya daha küçük olanlar ağır suç işlemektedir.

KURAL 2.5 (15, 0,733): Akdeniz Bölgesi doğumlu, serbest meslek sahibi ve yaşı 19' dan büyük olanlar ağır suç işlemektedir.

KURAL 2.6 (21, 0,714): Doğum yeri bilinmeyen veya Akdeniz Bölgesi olan, yaşı 19' dan büyük fakat 28 veya daha küçük olan ve okur yazar olanlar ağır suç işlemektedirler.

KURAL 2.7 (7, 0,714): Doğum yeri bilinmeyen, ev hanımı ve yaşı 19' dan büyük olanlar ağır suç işlemektedir.

KURAL 2.8 (10, 0,7): Emekli olan ve doğum yeri bilinmeyen veya Akdeniz Bölgesi olanlar ağır suç işlemektedir.

KURAL 2.9 (56, 0,643): Yaşı 18' den büyük fakat 22 veya daha küçük olan, okur yazar ve erkek olanlar ağır suç işlemektedir.

KURAL 2.10 (6.319, 0,592): Doğu Anadolu, Ege, Güney Doğu Anadolu, İç Anadolu, Karadeniz, Marmara Bölgesi doğumlu veya göçmen olanlar ağır suç işlemektedir.

Üçüncü kümeyi oluşturan ve (\$KX-Kohonen=3, \$KY-Kohonen=0) kümesinde yer alan bireyler için C5.0 algoritması yardımıyla türetilen 12 adet kural ise aşağıdaki gibidir:

KURAL 3.1 (14, 1,0): Ege veya İç Anadolu Bölgesi doğumlu, esnaf ve sanatkar olan ve yaşı 40' ın üzerinde olanlar ağır suç işlemektedir.

KURAL 3.2 (10, 0,9): Ege veya İç Anadolu Bölgesi doğumlu, yaşı 20' den büyük fakat 22 veya daha küçük olanlar, ilkokul mezunu ve serbest meslek sahibi olanlar ağır suç işlemektedir.

KURAL 3.3 (3, 1,0): Üniversite mezunu, Ege veya İç Anadolu Bölgesi doğumlu ve yaşı 49 veya daha büyük olanlar ağır suç işlemektedirler.

KURAL 3.4 (54, 0,796): Yaşı 13' den büyük fakat 17 veya daha küçük olan, kadın ve Doğu Anadolu, Ege, Güney Doğu Anadolu, İç Anadolu, Karadeniz, Marmara Bölgesi doğumlu veya göçmen olanlar ağır suç işlemektedir.

KURAL 3.5 (7, 0,857): Ege Bölgesi doğumlu ve esnaf ve sanatkar olanlar ağır suç işlemektedir.

KURAL 3.6 (13, 0,769): Öğrenci, çiftçi veya emekli olan, Ege veya İç Anadolu Bölgesi doğumlu ve yaşı 17' den büyük olanlar ağır suç işlemektedirler.

KURAL 3.7 (365, 0,71): Ev hanımı, ilkokul mezunu ve Doğu Anadolu, Ege, Güney Doğu Anadolu, İç Anadolu, Karadeniz, Marmara Bölgesi doğumlu veya göçmen olanlar ağır suç işlemektedir.

KURAL 3.8 (1.161, 0,693): Orta okul mezunu, yaşı 17' den büyük ve Doğu Anadolu, Ege, Güney Doğu Anadolu, İç Anadolu, Karadeniz, Marmara Bölgesi doğumlu veya göçmen olanlar ağır suç işlemektedir.

KURAL 3.9 (1.550, 0,674): Lise veya meslek yüksek okulu mezunu, Doğu Anadolu, Ege, Güney Doğu Anadolu, İç Anadolu, Karadeniz, Marmara Bölgesi doğumlu veya göçmen ve yaşı 17' den büyük olanlar ağır suç işlemektedir.

KURAL 3.10 (2.036, 0,655): İşçi, yaşı 39 veya daha küçük fakat 17' den büyük olan, Doğu Anadolu, Ege, Güney Doğu Anadolu, İç Anadolu, Karadeniz, Marmara Bölgesi doğumlu veya göçmen ve erkek olanlar ağır suç işlemektedir.

KURAL 3.11 (11, 0,636): Ege veya İç Anadolu Bölgesi doğumlu, yaşı 46' dan büyük ve işçi olanlar ağır suç işlemektedir.

KURAL 3.12 (13.662, 0,613): Doğu Anadolu, Güney Doğu Anadolu, Karadeniz, Marmara Bölgesi doğumlu veya göçmen olan ve yaşı 17' den daha büyük olanlar ağır suç işlemektedirler.

Dördüncü kümeyi oluşturan ve (\$KX-Kohonen=1, \$KY-Kohonen=1) kümesinde yer alan ağır suç işleyen bireyler için C5.0 algoritması uygulanması sonucunda elde edilen 10 kural aşağıda verilmiştir:

KURAL 4.1 (20, 0,85): İç Anadolu Bölgesi doğumlu, okur yazar, yaşı 37 veya daha küçük, erkek ve olayı (00.00-05.59), (06.00-11.59) veya (12.00-17.59) arasında işleyenler ağır suç işlemektedir.

KURAL 4.2 (3, 1,0): İç Anadolu Bölgesi doğumlu, ilkokul mezunu, kadın ve yaşı 20' den büyük olanlar ağır suç işlemektedir.

KURAL 4.3 (10, 0,8): Ege Bölgesi doğumlu ve yaşı 20' den büyük fakat 29 veya daha küçük olanlar ağır suç işlemektedirler.

KURAL 4.4 (2, 1,0): Ege Bölgesi doğumlu ve okur yazar olmayanlar ağır suç işlemektedirler.

KURAL 4.5 (15, 0,733): Doğu Anadolu Bölgesi doğumlu, kadın, olayı (00.00-05.59) veya (12.00-17.59) arasında işleyen ve okur yazar olmayan veya okur yazar olanlar ağır suç işlemektedirler.

KURAL 4.6 (145, 0,683): Olayı (06.00-11.59) arasında işleyen, okur yazar olmayan veya ilkokul mezunu ve Doğu Anadolu, Ege, Güney Doğu Anadolu, İç Anadolu, Karadeniz, Marmara Bölgesi doğumlu veya göçmen olanlar ağır suç işlemektedir.

KURAL 4.7 (15, 0,667): Kadın, esnaf ve sanatkar ve Doğu Anadolu, Ege, Güney Doğu Anadolu, İç Anadolu, Karadeniz, Marmara Bölgesi doğumlu veya göçmen olanlar ağır suç işlemektedir.

KURAL 4.8 (1.078, 0,635): Marmara Bölgesi doğumlu, erkek, olayı (00.00-05.59), (06.00-11.59) veya (12.00-17.59) arasında işleyen ve yaşı 20' den büyük olanlar ağır suç işlemektedirler.

KURAL 4.9 (2.716, 0,543): Olayı (00.00-05.59), (06.00-11.59) veya (12.00-17.59) arasında işleyen bireyler ağır suç işlemektedir.

KURAL 4.10 (96, 0,542): Doğu Anadolu Bölgesi doğumlu, yaşı 20' den büyük fakat 38 veya daha küçük olan, okur yazar olmayan veya okur yazar, olayı (00.00-05.59), veya (12.00-17.59) arasında işleyen mesleği bilinmeyen, öğrenci, ev hanımı, esnaf ve sanatkar veya yönetici olan bireyler ağır suç işlemektedirler.

Son olarak beşinci kümeyi oluşturan ve (\$KX-Kohonen=1, \$KY-Kohonen=2) değerlerini sağlayan bireyler için C5.0 algoritması yardımıyla türetilen 14 adet kural ise aşağıdaki gibidir:

KURAL 5.1 (4, 1,0): Kadın, mesleği esnaf ve sanatkarlık olan, ilkokul mezunu ve yaşı 35' den büyük olan bireyler ağır suç işlemektedirler.

KURAL 5.2 (15, 0,867): Kamu personeli veya emekli olan ve Akdeniz bölgesi doğumlu olan bireyler ağır suç işlemektedirler.

KURAL 5.3 (26, 0,808): Yaşı 16' dan büyük fakat 19 veya daha az olan, lise mezunu, Güney Doğu Anadolu, Karadeniz veya Marmara Bölgesi doğumlu ve mesleği bilinmeyen, boşta gezen, öğrenci, çiftçi, Kamu personeli, serbest meslek sahibi, esnaf ve sanatkar, yönetici veya emekli olanlar ağır suç işlemektedirler.

KURAL 5.4 (7, 0,857): Meslek yüksek okulu veya üniversite mezunu, kadın, serbest meslek sahibi veya esnaf ve sanatkar olan ve Doğu Anadolu, Ege, Güney Doğu Anadolu, İç Anadolu, Karadeniz veya Marmara Bölgesi doğumlu olan bireyler ağır suç işlemektedir.

KURAL 5.5 (2, 1,0): 18 yaşında, okur yazar olmayan ve Marmara bölgesi doğumlu olan bireyler ağır suç işlemektedirler.

KURAL 5.6 (2, 1,0): 60 yaşında, esnaf ve sanatkar ve Marmara Bölgesi doğumlu olan bireyler ağır suç işlemektedirler.

KURAL 5.7 (8, 0,75): Kadın, Karadeniz bölgesi doğumlu, okur yazar, serbest meslek sahibi veya esnaf ve sanatkar olan ve yaşı 19' dan büyük olan bireyler ağır suç işlemektedir.

KURAL 5.8 (24, 0,708): Yaşı 21' den büyük fakat 24 veya daha küçük olan, Güney Doğu Anadolu veya Karadeniz Bölgesi doğumlu, okur yazar, serbest meslek sahibi ve erkek olan bireyler ağır suç işlemektedirler.

KURAL 5.9 (595, 0,676): Ortaokul mezunu, Güney Doğu Anadolu, Karadeniz veya Marmara Bölgesi doğumlu olan ve 16 yaşından büyük olan bireyler ağır suç işlemektedirler.

KURAL 5.10. (4, 0,75): 19 yaşında veya daha küçük olan, okur yazar olmayan ve Karadeniz Bölgesi doğumlu olan bireyler ağır suç işlemektedirler.

KURAL 5.11 (7, 0,714): 59 yaşından büyük olan ve Güney Doğu Anadolu Bölgesi doğumlu olan bireyler ağır suç işlemektedirler.

KURAL 5.12 (15, 0,667): Yaşı 16' dan büyük fakat 19 veya daha küçük olan, ev hanımı veya esnaf ve sanatkar olan, ilkokul mezunu, Güney Doğu Anadolu, Karadeniz veya Marmara Bölgesi doğumlu olan bireyler ağır suç işlemektedirler.

KURAL 5.13 (12, 0,667): Yaşı 18 olan, işçi, ilkokul mezunu, Doğu Anadolu, Ege, Güney Doğu Anadolu, İç Anadolu, Karadeniz veya Marmara Bölgesi doğumlu olan bireyler ağır suç işlemektedirler.

KURAL 5.14 (10.471, 0,56): Doğu Anadolu, Ege, Güney Doğu Anadolu, İç Anadolu, Karadeniz veya Marmara Bölgesi doğumlu olan bireyler ağır suç işlemektedirler.

#### IV. Sonuç

Kümeleme analizinde kullanılan yeni bir yaklaşım olan SOM, parametrik olmayışı nedeniyle popüler bir kümeleme analizi tekniği olarak karşımıza çıkmaktadır. Ağır suç işleyen bireylere ilişkin olarak uygulanan SOM analizi sonucunda, elde edilen kümelerin özelliklerinin araştırılmasında C5.0 karar algoritmasından yararlanılmıştır. C5.0 karar algoritması uygulanması sonucunda, ağır suç işleyen bireyler için 5 kümenin söz konusu olduğu anlaşılmıştır.

Birinci küme için elde edilen üç kurala bakıldığında, bu küme için ağır suç işlemede en önemli etkenin olay saati olduğu anlaşılabilir. Birinci kümeyi oluşturan bireyler olayı (00.00-05.59), (06.00-11.59) veya (18.00-23.59) saat dilimlerinde işlemektedirler.

İkinci küme için bulunan 10 kural incelendiğinde, bulunan kuralların çoğunluğunda kuralları sağlayan bireylerin ortak özelliğinin 19 yaşından büyük olmaları ve doğum yerinin bilinmemesi veya Akdeniz Bölgesi olması olduğu anlaşılabilir.

Üçüncü küme için elde edilen 12 kural incelendiğinde ve bulunan kuralların ortak yönlerine bakıldığında, Ege veya İç Anadolu Bölgesi doğumlu olan ve 17 yaşından büyük bireylerin bu üçüncü küme içerisinde yer aldığı anlaşılabilir.

Dördüncü kümenin özelliklerini gösteren 10 kuralın ortak özellikleri arasında, bu kümede yer alan bireylerin olayı çoğunlukla (00.00-05.59) veya (12.00-17.59) arasında işlemeleri, İç Anadolu veya Ege Bölgesi doğumlu ve 20 yaşından büyük olmaları sayılabilir.

Beşinci ve son küme için elde edilen 14 kurala topluca bakıldığında ise, bu kümede yer alan bireylerin Güney Anadolu, Karadeniz veya Marmara Bölgesi doğumlu olmaları ve esnaf ve sanatkar olmaları ortak özellikleri olarak belirtilebilir. Kümelere ilişkin bu genel yorumların yapılmasında örnek sayıları ve güven değerleri dikkate alınmıştır.

#### Kaynaklar

Clementine 7.0 User's Guide (2002), Integral Solutions Limited, USA.  
Dasu Tamraparni ve Johnson Theodore (2003), Exploratory Data Mining and Data Cleaning, John Wiley&Sons, USA.

- Dunham Margaret H. (2003), *Data Mining Introductory and Advanced Topics*, Prentice Hall, USA.
- Flexer Arthur (2001), "On the Use of Self-Organizing Maps for Clustering and Visualization", *Intelligent Data Analysis* 5(2001).
- Kiang Melody Y (2001), "Extending the Kohonen Self-organizing Map Networks for Clustering Analysis", *Computational Statistics&Data Analysis*, 38(2001).
- Kiang Melody Y. ve Kumar Ajith (2001), "An Evaluation of Self Organizing Map Networks as a Robust Alternative to Factor Analysis in Data Mining Applications", *Information Systems Research*, Vol.12, No.2.
- Kloptchenko A., Eklund T., Karlsson J., Back B., Vanharanta H Ve Vısa A. (2004), "Combining Data and Text Mining Techniques for Analysing Financial Reports", *Intelligent Systems In Accounting, Finance and Management*, Jan-Mar., 12, 1.
- Wang Shouhong (2001), "Cluster Analysis Using a Validated Self-organizing Method: Cases of Problem Identification", *International Journal of Intelligent Systems in Accounting, Finance&Management*, 10 (2001).
- Wang Shouhong ve Wang Hai (2002), "Knowledge Discovery Through Self-Organizing Maps: Data Visualization and Query Processing", *Knowledge and Information Systems*, 4.
- Oğuzlar Ayşe (2004), *Veri Madenciliğine Giriş*, Ekin Kitabevi, Bursa.
- Penn Brian S. (2005), "Using Self-Organizing Maps to Visualize High-dimensional data", *Computers&Geosciences*, article in press.