
KAÇAKÇILIKTA YAKALANAN MALZEME TÜRLERİNE GÖRE SUÇLULARIN KÜMELENMESİ: İKİLİ KÜMELEME YÖNTEMİ

Ramazan ARSLAN¹, H. Hasan ÖRKÜ², Bülent ALTUNKAYNAK³

Öz

Kaçakçılık, ülkemizi ekonomik ve sosyal yönden olumsuz etkileyen, aynı zamanda terör ve suç örgütlerinin faaliyetlerini sürdürebilmesi için finansal destek sağlayan çok boyutlu, çok aktörlü ve değişken yapılı bir süreçtir. Suçlara karşı önlemlerin alınmasında suçlu profillerinin incelenmesi de ayrıca kritik bir öneme sahiptir. Bu çalışmanın amacı kaçakçılıkla ilgili suçluların yakalandıkları malzemelere göre kümelenebilir. Bu sayede birbiriyle ilişkili kaçakçılık malzemelerine ilişkin kümeler oluşturulmuş ve bu kümelere ait suçluların ortak özellikleri belirlenmiştir. Bu çalışmada hem kaçakçılık malzemeleri hem de suçlular aynı anda kümelendikleri için klasik kümeleme yöntemlerinden farklı olarak ikili kümeleme yönteminin kullanımı önerilmektedir. İkili kümeleme yöntemi olarak Bimax algoritması kullanılmıştır. Elde edilen sonuçlara göre, birbiriyle ilişkili malzemeler kümelenebilir ve bu kümelere yer alan suçluların bir araya getirilmesiyle de suçlu profillerine ilişkin bir takım özellikler ortaya çıkarılmıştır.

Anahtar Kelimeler: İkili kümeleme, Bimax, Suç verisi, Yakalanan malzeme, Suçlu profili.

JEL Sınıflandırması: C18, C38, C89.

CLUSTERING OF CRIMINALS ACCORDING TO THE TYPES OF MATERIAL CAUGHT IN SMUGGLING: BICLUSTERING METHOD

Abstract

Smuggling is a multidimensional, multi-actor and variable structure process that negatively affects to our country in the economic and social aspects and same time provides financial support for terror and criminal organizations to carry on their activities. Investigation of criminal profiles in taking measures against crimes also has a critical importance. The aim of this study is to clusters of smuggling-related criminals according to the material they are caught. In this way, clusters of related trafficking materials were established and common characteristics of the criminals belonging to these clusters were determined. In this study, the use of the bicluster method is proposed as it is different from classical clustering methods because both trafficking materials and criminals are clustered at the same time. Bimax algorithm is used as binary clustering method. According to the results obtained, the related materials were clustered and by bringing together the criminals in these clusters, a number of characteristics related to the criminal profiler were revealed.

Keywords: Biclustering, Bimax, Crime data, Captured material, Criminal profile.

JEL Classification: C18, C38, C89.

¹ Doktora Öğrencisi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Bölümü, ra325@mynet.com

² Doç. Dr., Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, bulenta@gazi.edu.tr

³ Doç. Dr., Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, hhorkcu@gazi.edu.tr

1. Giriş

Suçlar sosyal, siyasal, ekonomik, fiziksel ve psikolojik şartlar ile coğrafi faktörler arasındaki etkileşimden kaynaklanan bir olgudur. Suçların nedenlerinin tespiti, önlenmesi ve mücadele edilebilmesi için suçun meydana geldiği alanların ve mekânsal özelliklerin bilinmesi gerekmektedir. Suç eyleminden sorumlu bireylerin kişilik özelliklerini gösteren yöntem suçlu profili olarak tanımlanmaktadır. Suçlu profili çıkarmanın amacı, şüphelileri bulmaya yardımcı olabilecek kişilik bilgilerini sağlamak, davranış eğilimlerini ve demografik karakteristiklerini ortaya koymaktır.

Bu çalışmada Türkiye’de kırsal alanda kaçakçılık suçundan yakalanan suçluların yakalandıkları malzemeler dikkate alınmaktadır. Kaçakçılık suçlularının yakalandıkları malzemelere göre kümelenmesi ve bu suçlu kümelerinin demografik özelliklerinin incelenmesi birbiriyle ilişkili kaçakçılık malzemelerinin takibinde ve suçluların tespit edilmesinde önemlidir.

Suç verileri nispeten büyük hacimli verilerdir. Yargıtay Cumhuriyet Başsavcılığı’na 2007 yılında gelen toplam dosya sayısı (geçen yıldan devir ve yıl içinde gelen toplam dosya) 531 660 iken, %63.7’ lik artış oranı ile 2016 yılında 870 225’ e ulaşmıştır. (Adalet Bakanlığı Adli Sicil ve İstatistik Genel Müdürlüğü, 17 Mayıs 2017). Bu nedenle suç verilerinin analizinde özellikle kümelemeye dayalı veri madenciliği teknikleri yaygın olarak kullanılmaktadır.

Brown (1998:2848-2853) veri madenciliği tekniklerini kullanarak suçları mekânsal, zamansal ve niteliksel açıdan ilişkilendirmiştir.

Oğuzlar (2005:95-107) Self Organizing Map algoritmasına dayalı olarak ağır suç işleyen bireyleri 5 kümeye ayırmışlardır.

Bruin ve vd. (2006:171-177) suçlular ve suç davranışlarını tanımlamak üzere suçun ciddiyeti, sıklığı, ortamı ve süresi gibi faktörleri veri madenciliği yöntemleriyle incelemiştir. Bu sayede suçlara ilişkin görsel kümeleri oluşturularak suçluların ortak özelliklerinin belirlenmesi sağlanmıştır.

Nath (2006:41-44) suç örüntülerini (pattern) tespit etmek için kümelemeye dayalı veri madenciliği yöntemlerini kullanmıştır. Çalışmada, k-means algoritması kullanılarak suç örüntüsünün yapısı tespit edilmiş ve mekânsal haritalar yardımıyla elde edilen sonuçlar görselleştirilmiştir.

Tüzüntürk (2009:71-91) Türkiye’de 2006 yılı suç verilerine Principal Component Analysis uygulamış ve suç türlerini bir araya getirerek ilişkilendirmiştir.

Ma vd. (2010:218-223) büyük veri kümelerinden benzer olay alt kümelerini bulmak için iki adımlı bir kümeleme algoritması (AK-Modes) önermişlerdir. Veriler; Çin’in 11 farklı eyaletine ait şehirlerde meydana gelen 2004-2010 yılları arasındaki hırsızlık olaylarını kapsamaktadır. Deney verileri, 16 davranışsal özellik taşıyan 18 428 hırsızlık vakası kayıtlarını içermektedir. Deney verileri eğitim ve test verileri olarak ikiye bölünmüştür. Eğitim verileri, istihbarat analistlerinin tavsiyesine göre, meydana gelen bölge, konum kategorisi, mağdur / hedef, gasp yolları, modus operandi (harekât tarzı), suç motivasyonu, suç karakteristiği ve bir suçu işlemek için uygun durum gibi yalnızca 8 özel nitelik seçilmiştir. Sonuç olarak; 10 farklı guruba ayrılmış toplam 489 farklı vaka tespit edilmiştir.

Polat vd. (2013:1-33), Türkiye’de ilçe düzeyinde hırsızlık olaylarını analiz etmişlerdir. Çalışma kapsamında 2012 yılı için ilçe seviyesinde suç yoğunluğu ortaya konmuş ve hırsızlık suçunu etkileyen sosyo-ekonomik faktörler belirlenmiştir.

Giray (2016:1-31), bir cezaevinin hükümlü kütüğü verilerinden yararlanarak, benzer özellikler gösteren hükümlülere ait alt grupların değerlendirmesini iki aşamalı kümeleme analizi ile gerçekleştirmiştir.

Sea vd. (2016:3-21), Güney Kore'deki 111 cinsel saldırı davasından 16 farklı suç türü belirlemiş ve elde ettiği veri setine hiyerarşik kümeleme analizini uygulamıştır. Elde edilen sonuçlar, cinsel saldırılarda 6 kümeye ya da davranış profiline dikkat çekmektedir.

Reale vd. (2017:120-133), Kanada'da cinsel cinayet suçlarını işleyenlerin bir örneğinde farklı sadist suçları tanımlamak için çalışmış ve bu amaç için hiyerarşik kümeleme analizini kullanmıştır.

Yapılan çalışmalardan da görüldüğü gibi suç verilerinin analizinde kümeleme yöntemleri yaygın olarak kullanılmaktadır. Ancak, yukarıda verilen örneklerden de görüldüğü gibi klasik kümeleme yöntemleri veriyi tek bir boyutta (suçluların benzer özelliklere göre kümeleneceği, suç türlerinin ilişkilendirilmesi gibi) ele alarak kümeleme işlemini gerçekleştirmektedir. Oysa bir veri yapısını iki boyut (satır ve sütunlar) açısından dikkate alarak kümelemek suçların incelenmesinde daha derinlemesine bir inceleme sağlayacaktır. İki boyutu dikkate alan kümeleme yöntemlerine biclustering yöntemleri adı verilmektedir (Hartigan 1972:123-129; Cheng ve Church, 2000:93-103).

Türkiye'de en fazla görünen suç türlerinden birisi de kaçakçılıktır. Otoritelerin belirlediği vergiden kaçınmak, kısıtlı ya da yasak ürünleri toplumun kullanımına sunmak ve bunun için farklı yöntemler geliştirip haksız kazancın oluşumunu sağlamak geçmişte ve günümüzde kaçakçılığın gözle görülebilir tipik ortak değerleridir. Yüksek miktarlarda haksız kazanç elde etmenin yollarından biri olan kaçakçılık, toplumun fertleri arasındaki fırsat ve dolayısıyla rekabet eşitliğine zarar vermekte ve sosyal adalet dengesini de sarsmaktadır. Bunda kaçakçılığı bir yaşam tarzı olarak gören, kolay para kazanma arzusu içerisinde olan ve alt kültürde yetişen bireylerde sosyal statü elde etme eksikliğinin etkisi vardır. Türkiye'de en fazla görülen kaçakçılık malzemeleri sigara, akaryakıt, çay, cep telefonu ve uyuşturucu maddedir (Emniyet Genel Müdürlüğü KOM Daire Başkanlığı, 2015). Diğer yandan, tek bir olayda birden fazla kaçakçılık malzemesi de ele geçirebilmektedir. Örneğin, sim kart ve cep telefonunun veya silah ve uyuşturucunun bir arada ele geçirilmesi gibi. Buna ilaveten ele geçirilen malzeme türlerine göre suçluların sosyo-demografik özellikleri farklılık gösterebilmektedir. Örneğin cep telefonu kaçakçılığında yakalananların genellikle Suriye uyruklu olması gibi. Bu nedenle birbiriyle ilişkili malzemelerin kümeleneceği ve bu kümede yer alan suçluların bir araya getirilerek suçlu profillerine ilişkin desenlerin ortaya çıkarılması ile suç bağlantılarının bulunması suçla mücadele açısından önemlidir.

Bu çalışmada, Türkiye 2014 yılı kaçakçılık suç verisi dikkate alınmıştır. Çalışmanın temel amacı suçluları ve suç malzemelerini aynı anda kümeleyerek suçlu kümelerine ilişkin profillerin oluşturulmasıdır. Çalışmanın 2. bölümünde biclustering yöntemi tanıtılmış ve bu çalışmada kullanılan Bimax algoritmasının ayrıntıları verilmiştir. 3. bölümde veri setine ilişkin açıklamalar yer almaktadır. Bu bölümde ayrıca çalışmada kullanılacak veri matrisinin yapısı ve kullanılan istatistiksel testler hakkında bilgi verilmiştir. 4. bölüm elde edilen analiz sonuçlarını içermektedir. Bu bölümde biclustering algoritmasıyla elde edilen sonuçlar tablolar ve grafikler halinde verilerek ayrıntılı bir şekilde yorumlanmıştır. 5. bölüm bir önceki bölümde elde edilen sonuçlara ilişkin tartışmaları içermektedir. Son bölümde ise çalışmanın sonucu verilmiştir.

2.Yöntem

2.1. İkili Kümeleme (Biclustering)

Biclustering yöntemi literatürde matris kümelemesi, co-clustering veya blok kümeleme yöntemi olarak da bilinmektedir (Govaert ve Nadif, 2008:3233-3245; 2013; Van Mechelen vd., 2004:363-394). Bu kavram ilk defa Hartigan (1972:123-129) tarafından kullanılmasına rağmen yaygınlaşması Cheng ve Church (2000:93-103)'ün çalışmasıyla başlamıştır. Biclustering yönteminin temel mantığı veri matrisine dayalı olarak satır ve sütunları aynı anda kümelemektir (Şekil 1). Bu yönü ile sadece satır veya sadece sütun kümelemesi yapan klasik kümeleme yaklaşımlarından farklıdır. Ancak bu farklılık hesaplama zorluğunu da beraberinde getirmektedir (Zhao vd.: 2007-264-274, 2009:2578-2588). Bu nedenle biclustering tekniğinin yaygınlaşması bilgisayar gelişimiyle doğru orantılı olmuştur. Özellikle son yıllarda microdizi verilerin analizi için biclustering teknikleri yaygın olarak kullanılmaya başlanmıştır. Bunun nedeni microdizi verilerde satırları gösteren genlerle sütunları

gösteren koşulların birlikte kümelenmesinin gerekliliğidir. Bu nedenle biclustering yöntemine dayalı uygulamaların çoğu gene ifade verileriyle ilgilidir. Bu verilerin satır ve sütun sayısı bakımından çok büyük bir yapıya sahip olması Cheng ve Church (2000:93-103)'ün önerdiği CC algoritmasının dışında farklı algoritmaların gelişmesine de neden olmuştur. Bunlara örnek olarak, Prelic vd. (2006:1122-1129) tarafından geliştirilen Bimax algoritması, Lazzeroni ve Owen (2002:1-26) tarafından geliştirilen ve sonradan Turner vd. (2003:235-254) tarafından iyileştirilen The Plaid Model Bicluster algoritması, Murali ve Kasif (2003:78-88) tarafından geliştirilen The Questmotif Bicluster algoritması ve The Xmotifs Bicluster algoritması ve Kluger vd. (2003:703-716) tarafından geliştirilen The Spectral Bicluster algoritması verilebilir.

Bicluster algoritmalarının gen ifade verilerinin dışındaki farklı veri yapılarında da uygulandığı görülmektedir. Örneğin, bu alanda bir ilk olarak kabul edilen çalışma Hartigan (1972:123-129) tarafından ele alınmıştır. Çalışmada, iki farklı veri grubuna doğrudan kümeleme yöntemi uygulanmıştır. İlkinde, 1900-1968 yılları arasında güney eyaletlerinde Birleşik Devletler Başkanlığı için, Cumhuriyetçi oy oranından oluşan oylama verilerini kullanarak; aynı şekilde oy kullanan eyalet gruplarının ve oyların benzer olduğu yılların tespiti, ikincisinde ise 1969-1970 yıllarında Birleşmiş Milletler oylarından oluşan oylama verilerini kullanarak aynı çıkarlara veya siyasi sistemlere sahip ülkelerin ve önermelerin kümelerinin tanımlaması ve aynı temel konular hakkında önermeler dizisinin belirlenmesi amaçlamıştır.

Hoffman ve Puzicha (1999:668-693) internette yer alan ve yaklaşık üç milyon oydan oluşan EveryMovie veri kümesine Gibbs ikili kümelemesini uygulayarak internette benzer oy alan filmlerin kümelendiği çalışmalarını sunmuşlardır.

Lazzeroni ve Owen (2000:1-26), beslenme verilerine ve döviz örneğine ikili kümelemeyi uygulayarak benzer özelliklere sahip gıdaların alt kümelerini ve 1 ABD Doları için aynı miktarda alınabilecek benzer para birimlerini belirlemeye çalışmışlardır.

Meeds ve Roweis (2007:1-12), çalışmalarında bir veri matrisinin satır ve sütunlarını eş zamanlı olarak bir araya getiren bir olasılıksal blokla sabit bir ikili küme modeli sunmuşlardır. İkili küme parametrelerini analitik olarak bütünleştirmek için Gauss modeli kullanarak verimli bir örnekleme algoritması üretmeyi amaçlamışlardır.

Busygın vd. (2008:2964-2987), en yaygın kullanılan ve başarılı olan ikili kümeleme tekniklerini ve bunlarla ilgili uygulamaları incelemişlerdir. Çalışmaları, mevcut ikili kümeleme tekniklerinde karşılanabilir matematiksel kavramları vurgulayan teorik bir bakış açısı kazandırmıştır.

Wang vd. (2016:101-109) cep telefonu pazarında müşteri beğenilerini dikkate alan bicluster çalışması yapmışlardır. Bu çalışmada BCBimax algoritmasını kullanarak müşterilerin cep telefonu uygulamalarından şikâyetlerini dikkate alarak müşteri gruplarını tanımlamışlardır.

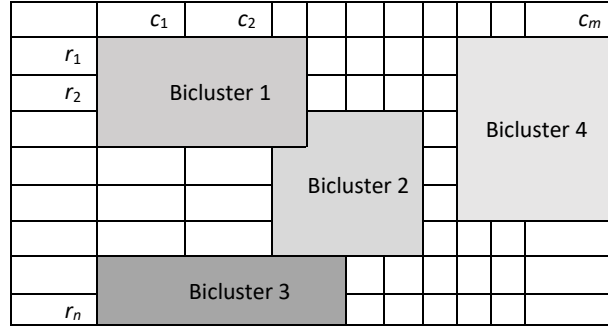
Raponi vd. (2016:31-45) 2009/2010 akademik yılında 55 İtalyan Ekonomi Fakültesini kapsayan ve farklı kaynaklardan toplanan fakülte düzeyindeki verileri inceleyerek, gözlemlenen üniversiteler arasındaki göstergeler ve benzerlikler arasındaki ilişkiyi hesaba katmak için model tabanlı bir biclustering yaklaşımı kullanmışlardır.

Denitto vd. (2017:114-124), ikili kümeleme için yeni bir kompakt Faktör Grafiği önermiştir. Çalışmalarında tek bir ikili küme için sıralı arama olarak maksimum toplam algoritmayı kullanarak verimli bir şekilde çözülebilen ikili ve kompakt bir faktör grafiği elde etmeyi amaçlamışlardır. Önerilen yöntem olumlu bir şekilde mevcut teknoloji ile karşılaştırılmıştır. Ayrıca bu yaklaşım ile dört veri kümesine ait veriler (ikisi yapay, ikisi gerçek veri) en yeni yöntemlerle test edilmiş ve karşılaştırılmış, hem faktör grafiklerine dayanan önceki yaklaşımlara hem de diğer son teknoloji yöntemlere göre cesaret verici sonuçlar vermiştir.

Veroneze vd. (2017:288-309), mevcut literatürde bulunan sadece sabit değerlere sahip maksimum değerli ikili kümeleri bulma problemine numaralandırma veya tüm ikili kümeleri bulamayan veya elde edilen ikili kümeleri maksimum olarak destekleyen bulgusal tabanlı

yaklaşımları geliştirmek için verimli çözümler sunmuşlardır. Bu çözüm için In-Close2 olarak adlandırılan etkili bir kavram analizi algoritması sunmuşlardır.

Şekil 1: İkili Kümelerin Gösterimi



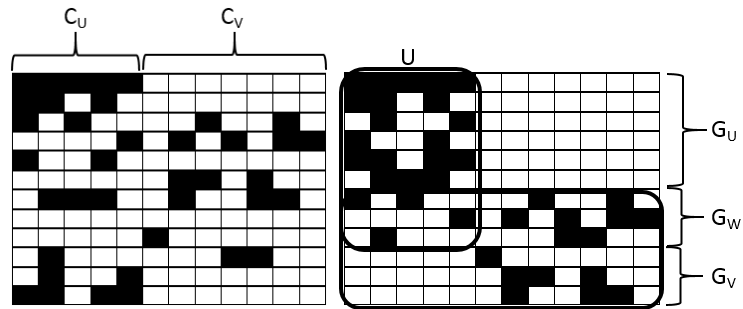
Bu çalışmada biclusterları elde etmek amacıyla Bimax algoritması uygulanmıştır. Bu algoritma, elemanları iki değer alan matrisler için en etkin algoritmadır (Prelc vd. 2006:1122-1129).

2.2. Bimax Algoritması

Prelc vd. (2006:1122-1129) tarafından geliştirilen Bimax algoritması, 0 ve 1 değerlerini alan bir matriste 1 değerinin alt gruplarını bulmak için kullanılan bir yöntemdir. Veriler farklı değerler alıyorsa veya değer aralıkları ilgi çekici ise ve iki değer alan matriste değerleri 1 ile işaretlemek mümkünse, Bimax algoritması yalnızca bu tür değerleri içeren alt grupları bulur. Prelc vd. (2006:1122-1129) farklı algoritmaları sabit bir kıyaslama için karşılaştırmada bu yöntemi kullanmış, ancak yöntem, iki değerli veya yarı iki değerli verilerin kullanıldığı diğer birçok uygulama alanında da fayda göstermiştir.

Bir veri matrisi, örneğin "binarize" işlevi kullanılarak iki değerli bir matrise dönüştürülebilir. Daha spesifik olarak, Şekil 2'de gösterilen Bimax algoritmasının ardındaki ana fikir, iki değerli matrisi, yalnızca biri 0 hücrelerini içeren üç alt matriste bölmektir, bunlardan yalnızca bir tanesi 0'dır ve bu nedenle gözardı edilebilir. Ardından, kalan iki alt matris U ve V'ye yinelenmeli olarak algoritma uygulanır; geçerli matris bir biclusteri temsil ediyorsa, yani sadece 1'leri içeriyorsa tekrarlama sona erdirilir. U ve V, iki değerli matrisin herhangi bir satırını ya da sütununu paylaşmıyorsa, yani GW boşsa, iki matris birbirinden bağımsız olarak işlenebilir. Bununla birlikte, eğer U ve V, Şekil 2'de gösterildiği gibi ortak bir GW sırasına sahiplerse, V'de sadece CV ile en az bir ortak sütun paylaşan bu biclusterların üretilmesi için özel bir dikkat gösterilmelidir. Çakışmaları önlemek için, bir sonraki bicluster, daha önce bulunan biclusterlerin satırları dışındaki verilere temel algoritmayı başlatarak bulunur (Prelc vd., 2006:1122-1129).

Şekil 2: Bimax Algoritmasının Gösterimi



Algoritmanın adımları aşağıdaki gibi verilebilir (Prelc vd., 2006:1122-1129).

Şekil 3: Bimax Algoritmasının Adımları

1. Divide the columns in two sets, C_U and C_V , based on the first row (in the first row, C_U contains only ones, while C_V contains only zeroes, see left panel in Fig. 2).
2. Resort the rows so that the first group (G_U) responds only in conditions given in C_U , the second group (G_W) responds both in C_U and C_V , and the last group (G_V) responds only in conditions given in C_V .
3. Define submatrices $U = C_U \cap (G_U \cup G_W)$ and $V = G_W \cup G_V$. The remaining part of the original matrix, $C_V \cap G_U$, contains only zeroes and is thus disregarded (the submatrix A in the right panel of Figure 2).
4. Repeat the previous steps for matrices U and V . If $G_W = \emptyset$, U and V can be treated independently from each other. Otherwise, the newly generated biclusters in V have to share at least one common column with C_V .

3. Veri Seti

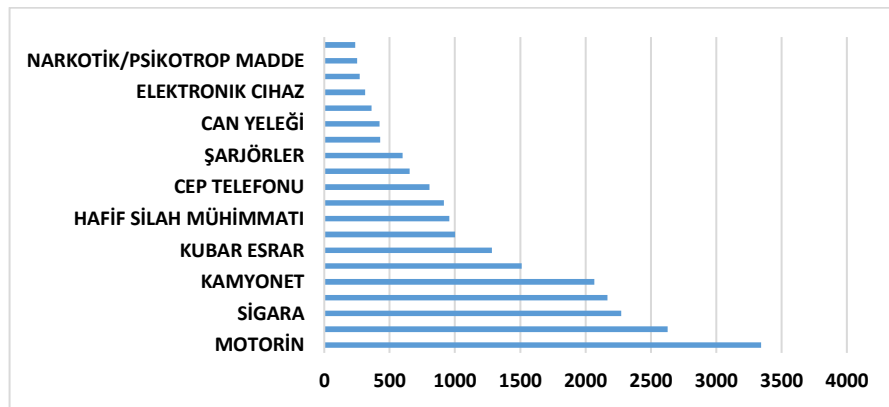
Çalışmanın veri seti 2014 yılına ait Jandarma Genel Komutanlığı Olaylar Bilgi Sisteminde toplanan kaçakçılıkla ilgili 41 164 suç kaydından oluşmaktadır. Veride suç türlerine, suç malzemelerine ve suçluların demografik özelliklerine (cinsiyet, medeni durum, öğrenim durumu, meslek, uyruk ve yaş) ait değişkenler yer almaktadır. Çalışmada, dörtten fazla farklı malzeme ile yakalanan suçlar dikkate alınmıştır. Bu suçluların sayısı 479'dur. Yıl içerisinde 50 den fazla kez yakalanan 22 çeşit malzeme dikkate alınmıştır. Bu malzemelerin kodları ve açıklamaları aşağıdaki tabloda verilmiştir.

Tablo 1: Suçlara İlişkin Kodlamalar

Malzeme Kodu	Malzeme açıklaması	Malzeme Kodu	Malzeme açıklaması
V3	Çay	V14	Motorin
V4	Diğer narkotik/psikotrop madde	V15	Muhtelif narkotik/psikotrop madde
V5	Eroin	V16	Muhtelif tütün mamülü
V6	Eski eser	V17	Narkotik/psikotrop madde
V7	Extacy	V18	Narkotik/psikotrop tabletler
V8	Haşhaş	V19	Para, paraya eşit sayılan değerler, kıymetli damga ve mühür
V9	Kenevir	V20	Patlayıcılar
V10	Kenevir tohumu	V21	Sigara
V11	Kokain	V22	Toz esrar
V12	Kubar esrar	V23	Tütün
V13	Metamfetamin	V24	Uzun namlulu silahlar

Bu malzemelere ilişkin dağılım Şekil 4'te verilmiştir.

Şekil 4: En Fazla Ele Geçirilen 10 Malzeme Türüne Ait Dağılım



Şekil 4'ten görüldüğü gibi en çok yakalanan üç malzeme sırasıyla motorin, otomobil ve sigaradır. Bu veri seti kullanılarak biclustering amaçlı veri matrisi oluşturulmuştur. Bu veri matrisinin satırları suçluları sütunları ise malzemeleri göstermektedir. Matrisin elemanları i . suçlunun j . malzemesiyle yakalanıp yakalanmadığını göstermektedir (Şekil 5). Dolayısıyla analiz için 479×22 boyutunda bir veri matrisi kullanılmıştır.

Şekil 5: Veri Matrisinin Yapısı

Suçlu	Malzeme türü			
	V3	V4	L	V24
01	1	0		0
02	0	1		1
N				
479	1	1		0

Veri matrisi kullanılarak Bimax algoritmasına dayalı biclustering işlemi R-project 3.2.2 bicluster library si kullanılarak gerçekleştirilmiştir (Şekil 6).

Şekil 6: Suç Veri Setinin İkili Kümelenmesi İçin R Kodu

```
data(Crime)
x<-as.matrix(Crime,479,22)
res<- biclust(x, method=BCBimax(), minr=15, minc=5, number=1000)
writeclust(res)
biclustmember(res, x)
```

4. Bulgular

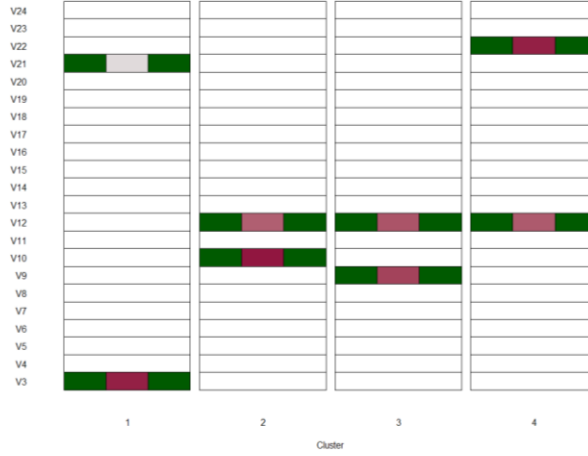
Suç verisi için Bimax algoritmasına göre elde edilen biclustering üyelikleri Tablo 2'de verilmiştir. Biclusterlar büyüklükleri bakımından incelendiğinde Bicluster 1'in 479 suçludan 56 suçluyu içerdiği (%11.69), Bicluster 2'nin 50 suçluyu içerdiği (%10.44), Bicluster 3'ün 309 suçluyu içerdiği (%64.51) ve Bicluster 4'ün 64 suçluyu içerdiği (%13.36) görülmektedir (Tablo 2).

Tablo 2: Bicluster Büyüklükleri

Bicluster	Büyüklük	Popülasyondaki oranı (%)
1	56 × 2	11.69
2	50 × 2	10.44
3	309 × 2	64.51
4	64 × 2	13.36

Bicluster üyelik grafiği Şekil 7'de verilmiştir.

Şekil 7. Bicluster Üyelik Grafiği



Burada her bir satır kaçakçılıkta yakalanan malzemeleri, her bir sütun ise bir ikili kümeyi göstermektedir. Yeşil bölümler yakalanan suçlular üzerinde ele geçen ortak malzemeleri açığa çıkarmaktadır. Bu yeşil alanların iç kısmında bulunan orta dikdörtgenler ise 0 (beyaz) ile 1 (kırmızı) arasında değişen, suçluların üzerinde yakalanan malzemelerin ortalama değerini belirtmektedir. Kırmızı alan ne kadar açık renk verirse, ikili kümelerdeki malzemeler ile yakalanan tüm malzemeler arasındaki uyum düzeyi de o kadar artmaktadır.

Şekil 7 ve Tablo 4 incelendiğinde aşağıdaki sonuçlara ulaşılabilir:

(1) Bicluster 1’de yer alan suçluların (56 suçlu, tüm suçluların %11,69’u) 5607 sayılı “Kaçakçılıkla Mücadele Kanununa Muhalefet” suçlarına ilişkin “Çay” (V3) ve “Sigara” (V21) malzemeleriyle yakalandığı görülmektedir. Orta dikdörtgenler dikkate alındığında bu küme için sigara yakalamalarının görülme oranının çay yakalamalarına göre daha fazla olduğu söylenebilir. Bu suçluların demografik özellikleri incelendiğinde ise suçluların büyük çoğunluğunun Van ve Mardin illerinde yakalandıkları (%56,2’si), Türk uyruklu oldukları (%73,7’si) ve Evli şahıslardan oluştuğu (%52,6’sı) görülmektedir.

(2) Bicluster 2’de yer alan suçluların (50 suçlu, tüm suçluların %10,44’ü) TCK Madde 188 gereği “Uyuşturucu veya Uyarıcı Madde İmal ve Ticareti” suçlarına ilişkin “Kenevir Tohumu” (V10) ve “Kubar Esrar” (V12) malzemeleriyle yakalandığı görülmektedir. Orta dikdörtgenler dikkate alındığında bu küme için kubar esrar yakalamalarının görülme oranının kenevir tohumu yakalamalarına göre daha fazla olduğu söylenebilir. Bu suçluların demografik özellikleri incelendiğinde ise suçluların büyük çoğunluğunun Bursa, İzmir ve Şanlıurfa illerinde yakalandıkları (%42’si), Türk uyruklu oldukları (%84’ü), Evli şahıslardan oluştuğu (%48’i) ve Eğitim durumlarının ilköğretim seviyesinde (%42’si) oldukları görülmektedir.

(3) Bicluster 3’te yer alan suçluların (309 suçlu, tüm suçluların %64,51’i) TCK Madde 188 gereği “Uyuşturucu veya Uyarıcı Madde İmal ve Ticareti” suçlarına ilişkin “Kenevir” (V9) ve “Kubar Esrar” (V12) malzemeleriyle yakalandığı görülmektedir. Orta dikdörtgenler dikkate alındığında bu küme için her iki yakalamanın da görülme oranlarının birbirlerine yakın oldukları söylenebilir. Bu suçluların demografik özellikleri incelendiğinde ise suçluların büyük çoğunluğunun Bursa ve İzmir illerinde yakalandıkları (%28,8’i), Türk uyruklu oldukları (%85,8’i), Evli şahıslardan oluştuğu (%54,2’i), Eğitim durumlarının ilköğretim seviyesinde (%49’u) ve Çiftçilerden oluştuğu (%33’ü) görülmektedir.

(4) Bicluster 4’te yer alan suçluların (64 suçlu, tüm suçluların %13,36’sı) TCK Madde 188 gereği “Uyuşturucu veya Uyarıcı Madde İmal ve Ticareti” suçlarına ilişkin “Kubar Esrar” (V12) ve “Toz Esrar” (V22) malzemeleriyle yakalandığı görülmektedir. Orta dikdörtgenler dikkate alındığında bu küme için kubar esrar yakalamalarının görülme oranının toz esrar yakalamalarına göre daha fazla

olduğu söylenebilir. Bu suçluların demografik özellikleri incelendiğinde ise suçluların büyük çoğunluğunun Çanakkale ve Şanlıurfa illerinde yakalandıkları (%38,4'ü), Türk uyruklu oldukları (%85'i), Bekâr şahıslardan oluştukları (%53,3'ü) ve Eğitim durumlarının ilköğretim seviyesinde (%21,7'si) ve Serbest Meslek sahiplerinden oluştuğu (%49'u) görülmektedir.

(5) Bicluster 2, 3 ve 4'te yer alan suçluların "Kubar Esrar" (V12) ile birlikte "Kenevir" (V9), "Kenevir Tohumu" (V10) ve "Toz Esrar" (V12) malzemeleriyle de yakalandıkları görülmektedir.

(6) Türkiye uyruklu suçlular tüm Bicluster'larda yoğunlaşırken, Suriye uyruklu suçlular düşük oranlarda Bicluster 1 ve 2'de görülmektedir.

(7) Evli olan suçluların ağırlıklı olarak Bicluster 1,2 ve 3'te yer aldığı bekârların ise Bicluster 4'te yer aldığı görülmektedir.

(8) Eğitim durumu sadece ilköğretim olan grubun Bicluster 2 ve 3'te yoğunlaştığı görülmektedir.

(9) Van ve Mardin'de yakalanan suçluların Bicluster 1'de, Bursa ve İzmir'de yakalanan suçluların Bicluster 2 ve 3'te, Çanakkale ve Şanlıurfa'da yakalanan suçluların ise Bicluster 4'te yoğunlaştığı görülmektedir.

(10) Mesleği Serbest olan suçluların tüm Bicluster'lerde yoğunlaştığı görülürken, mesleği Çiftçilik olan suçluların ise Bicluster 3'te yoğunlaştığı görülmektedir.

(11) Bicluster 3'teki suçluların yaş ortalamasının daha yüksek olduğu görülürken Bicluster 4'ün ise en genç yaş ortalamasına sahip bicluster olduğu görülmektedir.

Tablo 4: İkili Kümelerdeki Suçluların Demografik Özellikleri

Değişken	Değer	İkili Küme							
		1		2		3		4	
		f	%	f	%	f	%	f	%
Uyruk	Türkiye	42	73,7	42	84,0	247	85,8	51	85,0
	Suriye	6	10,5	0	0,0	0	0,0	4	6,7
	Diğer	9	15,8	8	16,0	41	14,2	5	8,3
Medeni durum	Bekâr	10	17,5	16	32,0	72	25,0	32	53,3
	Evli	30	52,6	24	48,0	156	54,2	17	28,3
	Bilinmiyor	17	29,8	10	20,0	60	20,8	11	18,3
Eğitim durumu	Okur-yazar	2	3,5	5	10,0	21	7,3	6	10,0
	İlköğretim	15	26,3	21	42,0	141	49,0	13	21,7
	Ortaöğretim	6	10,5	2	4,0	17	5,9	3	5,0
	Lisans ve üstü	6	10,5	3	6,0	13	4,5	11	18,3
İl	Bilinmiyor	28	49,1	19	38,0	96	33,3	27	45,0
	Van	18	31,6	0	0,0	0	0,0	0	0,0
	Mardin	14	24,6	0	0,0	3	1,0	0	0,0
	Bursa	0	0,0	9	18,0	38	13,2	0	0,0
	İzmir	1	1,8	6	12,0	45	15,6	0	0,0
	Çanakkale	0	0,0	1	2,0	1	0,3	13	21,7
	Şanlıurfa	0	0,0	6	12,0	8	2,8	10	16,7
Meslek	Diğer	24	42,1	28	56,0	193	67,0	37	61,7
	Serbest	13	22,8	13	26,0	56	19,4	16	26,7
	İşçi	6	10,5	0	0,0	10	3,5	2	3,3
	Çiftçi	4	7,0	10	20,0	95	33,0	4	6,7
	Diğer	4	7,0	6	12,0	31	10,8	6	10,0
	Bilinmiyor	30	52,6	21	42,0	96	33,3	32	53,3
Yaş	Ortalama±SS	32,42±10,32		40,00±13,67		41,00±14,55		29,48±9,46	

5. Sonuç

Kaçakçılık, ülkemizi ekonomik ve sosyal yönden olumsuz etkileyen, aynı zamanda terör ve suç örgütlerinin faaliyetlerini sürdürülebilmek için finansal destek sağlayan çok boyutlu, çok aktörlü ve değişken yapılı bir süreçtir. Kaçakçılık suçunda ele geçen malzeme türleri suçluların sosyo-demografik özelliklerine göre farklılık gösterebilmektedir. Bu nedenle birbiriyle ilişkili malzemelerin kümelenmesi ve bu kümede yer alan suçluların bir araya getirilerek suçlu profillerine ilişkin özelliklerin ortaya çıkarılması suçla mücadele açısından önemlidir.

Bu çalışmada Türkiye’de kaçakçılık suçundan yakalanan suçluların yakalandıkları malzemeler dikkate alınmaktadır. Kaçakçılık suçlularının yakalandıkları malzemelere göre kümelenmesi ve bu suçlu kümelerinin demografik özelliklerinin incelenmesi birbiriyle ilişkili kaçakçılık malzemelerinin takibinde ve suçluların tespit edilmesinde önemlidir. Bu amaçla hem suçluları hem de yakalanan malzeme türlerini dikkate alarak kümeleme yapan Bimax ikili kümeleme yöntemi ile Türkiye’de meydana gelen kaçakçılık suçlarında ele geçirilen malzemeler açısından suçlular kümelenmiştir. Çalışmamızda Bimax algoritması seçilmesinin nedeni geleneksel kümeleme yöntemlerine göre daha ayrıntılı ve gerçekçi sonuçlar sunmasıdır.

Türkiye’nin özellikle kıyı ve sınır bölgelerinde yaşayan yerel halk arasında kaçakçılık faaliyetlerine yönelik algının meşru bir eylem olduğu yönünde geliştiği görülmektedir. Kaçakçılığın doğal bir yaşam tarzı haline gelmesinin ötesinde yapılan bu eylemin bir suç olmadığı algısı da hâkim bir anlayıştır. Özellikle Doğu ve Güneydoğu sınırlarında yaşayan insanlar, kaçakçılıktan elde edilen gelirin bir kaçakçı için olağan, haklı ve helal bir kazanç olduğu düşüncesindedirler. Bu insanlar kaçakçılığın bir suç olmadığına inanmakta ve bunu tamamen bir ticaret ve geçimlerini temin aracı olarak algılamaktadırlar. Kaçakçılığın bir suç değil bir meslek olduğuna inanmaktadırlar.

Ülkemizin coğrafi yapısı, Asya, Orta Doğu ve Avrupa için sahip olduğu köprü pozisyonu, 5 000 milden fazla olan deniz sahili, 8 ülkeyle olan kara sınırı, gelir dağılımındaki adaletsizlik, işsizlik, büyük şehirlere göç, yozlaşmanın yaygın oluşu gibi faktörler, birçok suç faaliyetinin ortaya çıkmasını kolaylaştırmakta ya da sebep olmaktadır.

Elde edilen sonuçları incelediğimizde “Çay” (V3) ve “Sigara” (V21) malzemeleriyle beraber yakalanan suçluların birinci kümede, “Kenevir Tohumu” (V10) ve “Kubar Esrar” (V12) malzemeleriyle yakalanan suçluların ikinci kümede, “Uyuşturucu veya Uyarıcı Madde İmal ve Ticareti” suçlarına ilişkin “Kenevir” (V9) ve “Kubar Esrar” (V12) malzemeleriyle yakalanan suçluların üçüncü kümede ve “Uyuşturucu veya Uyarıcı Madde İmal ve Ticareti” suçlarına ilişkin “Kubar Esrar” (V12) ve “Toz Esrar” (V22) malzemeleriyle yakalanan suçluların ise dördüncü kümede oldukları görülmektedir.

İkinci, üçüncü ve dördüncü kümelerde yer alan suçluların “Uyuşturucu veya Uyarıcı Madde İmal ve Ticareti” suçlarına konu olan “Kenevir” (V9), “Kenevir Tohumu” (V10), “Kubar Esrar” (V12) ve “Toz Esrar” (V22) ile yakalandıkları yani bir bakıma ikinci, üçüncü ve dördüncü kümelerin tek bir küme olarak düşünülebileceği söylenebilir. Buradan çay ve sigara ile yakalananların bir kümede “Uyuşturucu veya Uyarıcı Madde İmal ve Ticareti” suçları ile yakalananların diğer kümede olduğu söylenebilir.

“Uyuşturucu veya Uyarıcı Madde İmal ve Ticareti” ile yakalanan suçluların büyük oranda bekâr, ilköğretim mezunu ve çiftçilik mesleği ile uğraşanlar oldukları gözlenmiştir.

Çay ve sigara kaçakçılığında yakalananların çoğunlukla Van ve Mardin illerinde, “Uyuşturucu veya Uyarıcı Madde İmal ve Ticareti” suçları ile yakalananların ise Bursa, İzmir, Çanakkale ve Şanlıurfa illerinde yakalandıkları görülmektedir.

Ülkemizin kara ve deniz sınırlarının uzunluğu dikkate alındığında, kıyı ve sınır kapılarımızda görev yapmakta olan Sahil Güvenlik Komutanlığı, Gümrük Teşkilatı, Polis ve Jandarma teşkilatlarının koordinasyon ve denetimlerinin uyum içerisinde olması kaçakçılıkla mücadelede

büyük önem taşımaktadır. Elde edilen sonuçların kullanılması, suçluların analiz edilmesi, suç önleme stratejileri geliştirmek için yardımcı olabilir.

Kaynakça

- Adalet Bakanlığı Adli Sicil ve İstatistik Genel Müdürlüğü (2017). *Haber Bülteni*, Sayı:11.
- Brown, D.E. (1998). The Regional Crime Analysis Program (RECAP): A Framework for Mining Data to Catch Criminals. *IEEE*, 2848-2853.
- Bruin, J.S., Cocx, T.K., Kusters, W.A., Laros, J. ve Kok J.N. (2006). Data Mining Approaches to Criminal Career Analysis. *In Proceedings of the Sixth International Conference on Data Mining (ICDM'06)*, 171-177.
- Busygin, S., Prokopyev, O. ve Pardalos, P.M. (2008). Biclustering in Data Mining. *Computers and Operations Research*, 35(9), 2964-2987.
- Cheng, Y. ve Church, G.M. (2000). Biclustering of Expression Data. *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*, 1, 93-103.
- Denitto, M., Farinelli, A., Figueiredo M.A.T. ve Bicego M. (2017). A Biclustering Approach Based on Factor Graphs and the Max-Sum Algorithm. *Pattern Recognition*, 62, 114-124.
- Emniyet Genel Müdürlüğü KOM Daire Başkanlığı, (2015). *2014 Kaçakçılık ve Organize Suçlarla Mücadele Raporu*, KOM Yayınları, Ankara.
- Giray, S. (2016). İki Aşamalı Kümeleme Analizi ile Hükümlü Verilerinin İncelenmesi. *Ekonometri ve İstatistik*, 25, 1-31.
- Govaert, G., ve Nadif, M. (2008). Block Clustering with Bernoulli Mixture Models: Comparison of Different Approaches. *Computational Statistics and Data Analysis*, 52 (6), 3233–3245.
- Govaert, G. ve Nadif, M. (2013). *Co-Clustering: Models, Algorithms and Applications*. ISTE, Wiley.
- Hartigan, J. A. (1972). Direct Clustering of a Data Matrix. *Journal of the American Statistical Association (JASA)*, 67(337), 123–129.
- Hofmann, T. ve Puzicha, J. (1999). Latent Class Models for Collaborative Filtering. *In Proceedings of the International Joint Conference on Artificial Intelligence*, 668–693.
- Kluger Y., Basri R., Chang J.T. ve Gerstein M. (2003). Spectral Biclustering of Microarray Data: Co-Clustering Genes and Conditions. *Genome Research*, 13, 703-716.
- Lazzeroni, L. ve Owen, A. (2000). Plaid Models for Gene Expression Data. *Technical Report, Stanford University*, 1-26.
- Ma, L., Chen, Y. ve Huang, H. (2010). AK-Modes: A Weighted Clustering Algorithm for Finding Similar Case Subsets. *IEEE*, 218-223.
- Meeds, E. ve Roweis, S. (2007). Nonparametric Bayesian Biclustering. *Department of Computer Science, University of Toronto*, 1-12.
- Murali, T. ve Kasif, S. (2003). Extracting Conserved Gene Expression Motifs from Gene Expression Data. *Pacific Symposium on Biocomputing*, 8, 77-88.
- Nath, S.V. (2006). Crime Pattern Detection Using Data Mining. *International Conferences on Web Intelligence and Intelligent Agent Technology-Workshops, IEEE/WIC/ACM*, 41-44.
- Oğuzlar, A. (2005). Kümeleme Analizinde Yeni Bir Yaklaşım: Kendini Düzenleyen Haritalar (Kohonen Ağları). *İktisadi ve İdari Bilimler Dergisi*, 19 (2), 93-107.
- Polat, C., Eren, H. ve Erbakıcı, F. (2013). Hırsızlık Suçunu Etkileyen Faktörlerin Değerlendirilmesi ve Geleceğe Yönelik Yaklaşımlar. *Güvenlik Bilimleri Dergisi*, 2(1), 1-33.

- Prelic, A., Bleuler, S., Zimmermann, P., Wil, A., Buhlmann, P., Grussem, W., Hennig, L., Thiele, L. ve Zitzler, E. (2006). A Systematic Comparison and Evaluation of Biclustering Methods for Gene Expression Data Bioinformatics. *Oxford Univ. Press*, 22, 1122-1129.
- Raponi, V., Martella, F. ve Maruotti, A. (2016). A Biclustering Approach to University Performances: An Italian Case Study. *Journal of Applied Statistics*, 43 (1), 31-45.
- Reale, K., Beauregard, E. ve Martineau, M. (2017). Sadism in Sexual Homicide Offenders: Identifying Distinct Groups. *Journal of Criminal Psychology*, 7(2), 120-133.
- Sea, J., Kim, K. ve Youngs, D. (2016). Behavioural Profiles and Offender Characteristics Across 111 Korean Sexual Assaults. *Journal of Investigative Psychology and Offender Profiling*, 13, 3-21.
- Turner, H., Bailey, T. ve Krzanowski, W. (2003). Improved Biclustering of Microarray Data Demonstrated Through Systematic Performance Tests. *Computational Statistics & Data Analysis*, 48 (2) ,235-254.
- Tüzüntürk, S. (2009). Çok Boyutlu Ölçekleme Analizi: Suç İstatistikleri Üzerine Bir Uygulama. *Uludağ Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Dergisi*, 28 (2), 71-91.
- Van Mechelen, I., Bock, H.H. ve De Boeck, P. (2004). Two-Mode Clustering Methods: A Structured Overview. *Statistical Methods in Medical Research*, 13 (5), 363–94.
- Veroneze, R., Banerjee, A., Fernando J. ve Zubena, V. (2017). Enumerating All Maximal Biclusters in Numerical Datasets. *Information Sciences*, 379, 288-309.
- Wang, B., Miao, Y., Zhao, H., Jing, J. ve Chen, Y. (2016). A Biclustering-Based Method for Market Segmentation Using Customer Pain Points. *Engineering Applications of Artificial Intelligence*, 47, 101–109.
- Zhao, H., Liew, A.W.C., Xie, X., ve Yan, H. (2007). A New Geometric Biclustering Algorithm Based on the Hough Transform for Analysis of Large-Scale Microarray Data. *J.Theor. Biol.* 251, 264–74.
- Zhao, H., Chan, K.L., Cheng, L.M., ve Hong, Y. (2009). A Probabilistic Relaxation Labeling Framework for Reducing the Noise Effect in Geometric Biclustering of Gene Expression Data. *Pattern Recognit*, 42 (11), 2578–2588.

CLUSTERING OF CRIMINALS ACCORDING TO THE TYPES OF MATERIAL CAUGHT IN SMUGGLING: BICLUSTERING METHOD

Extended Abstract

Aim: Smuggling is a multidimensional, multi-actor and variable structure process that negatively affects to our country in the economic and social aspects and same time provides financial support for terror and criminal organizations to carry on their activities. Investigation of criminal profiles in taking measures against crimes also has a critical importance. Since the traditional clustering methods used to construct crime profiles do clustering a single dimension, they gives general results than specific. One of the most challenging aspects of analyzing social science data is the presence of spatial effects in quantitative models of individual behavior. In this study, the use of the bicluster method is proposed as it is different from classical clustering methods because both trafficking materials and criminals are clustered at the same time. The aim of this study was to construct criminal profiles by clustering criminals and criminal materials at same time via the

biclustering method. By this method, criminal profiles will be constructed by clustering both the characteristics of the criminals and the types of material caught in smuggling at the same time.

Method(s): Biclustering is an important new technique in two way data analysis. After Cheng and Church (2000) followed the initial bicluster idea of Hartigan (1972) and started to calculate bicluster on microarray data, a wide range of different articles were published dealing with different kinds of algorithms and methods to preprocess and analyze the results of such methods. Biclustering techniques are capable of simultaneously clustering rows and columns of a data matrix. These techniques became very popular for the analysis of gene expression data, since a gene can take part of multiple biological pathways which in turn can be active only under specific experimental conditions. Several biclustering algorithms have been developed in the past recent years. Bimax algorithm is used as binary clustering method. The Bimax algorithm is a method that uses a data matrix consisting of values 0 and 1 and is shown as the most efficient algorithm in the existing binary clustering algorithms. The data set of the study consists of 86191 criminal records related to smuggling collected in the Gendarmerie General Command Events Information System in 2014. In the study, criminals caught with more than four different materials were taken into account. The number of these criminals is 447. Twenty kinds of materials caught more than 50 within the year were taken into consideration. Using this data set, a data matrix is composed for biclustering. The rows and columns of this data matrix show the criminals and the materials, respectively. The elements of the matrix show that whether or not the *i*th criminal was caught with *j*th material. 447x20 matrix of data was used in the study. The codes we used in our experiments were available as packages for the R language.

Findings: The three most frequently caught materials are diesel, car and cigarette. As a result of the Bimax algorithm, the criminals are divided into 3 clusters. In addition to the smuggling of mobile phones and sim card materials, the first group of criminals also made light weapon ammunition and short-barrel weapon smuggling. The second group of criminals are made smuggling short-barreled weapons and long-barreled weapons alongside smuggling of mobile phone and sim card materials. The criminals in the third group are seen to be possessed by weapons of smuggling. Turkish nationals are in the first cluster, Syrian nationals are in the second cluster and other nationalities are concentrated in the third group. The lowest group of education level seems to concentrate in the third group.

Conclusion: In this study, in terms of the materials seized in smuggling crimes in the Gendarmerie area of responsibility in Turkey, the criminals are clustered by the Bimax biclustering method. The implemented Bimax algorithm provides more detailed and realistic results than the conventional clustering methods. Using the results obtained can help to analyze the criminals and develop crime prevention strategies. According to the results obtained, the related materials were clustered and by bringing together the criminals in these clusters, a number of characteristics related to the criminal profiler were revealed.

