



SUÇ VERİLERİNİN İKİLİ KÜMELEME YÖNTEMLERİ İLE ANALİZİ

Demet ALBASAR

YÜKSEK LİSANS TEZİ
İSTATİSTİK ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

HAZİRAN 2025

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Demet ALBASAR

23/06/2025

SUÇ VERİLERİNİN İKİLİ KÜMELEME YÖNTEMLERİ İLE ANALİZİ
(Yüksek Lisans Tezi)

Demet ALBASAR

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Haziran 2025

ÖZET

Suç analizi, suçların öngörülebilirliği ve önleyici tedbirlerin alınması açısından güvenlik birimleri için kritik bir öneme sahiptir. Suç oranlarındaki artışla birlikte suç verileri, büyük ve karmaşık veri setleri haline gelmiş, bu verilerin verimli bir şekilde analiz edilmesi için veri madenciliği teknikleri ön plana çıkmıştır. Veri madenciliği, büyük veri kümelerinden gizli kalmış anlamlı desenleri ve ilişkileri keşfetmeye yardımcı olan bir yöntem olup suç verileri üzerinde de etkili bir şekilde kullanılmaktadır. Suç ve suçu meydana getiren faktörler bir arada incelenip, aralarındaki ilişkiyi tespit etmek ve çıkarımda bulunmak için pek çok amaçla suç analizi ile ilgili çalışma yapılmıştır. Bu analizler sayesinde, suç türleri doğru şekilde sınıflandırılabilir, suç eğilimleri belirlenebilir ve güvenlik önlemleri daha etkili bir şekilde planlanabilir. Özellikle suç bölgelerinin belirlenmesi gibi problemler için veri madenciliği algoritmaları suç türleri ve bu suçların işlendiği bölgeler arasındaki ilişkileri anlamada büyük bir rol oynamaktadır. Klasik kümeleme yöntemlerinde yalnızca satır veya sütun incelenip, aralarındaki ilişkiyi tespit etmek ve çıkarımda bulunmak için pek çok amaçla suç analizi ile ilgili çalışmalar yapılmıştır. Bu çalışmada, suç verilerinin analizinde ikili kümeleme yöntemleriyle suç bölgelerinin belirlenmesi amaçlanmıştır. Adalet Bakanlığı Adli Sicil ve İstatistik Genel Müdürlüğü'nden temin edilen veri seti üzerinde; R Project yazılımı kullanılarak CC, xMotif ve Bimax algoritmaları ile analizler gerçekleştirilmiştir. Algoritmaların ürettiği sonuçlar görsel açıdan karşılaştırılmıştır. Elde edilen sonuçlar, suç verilerinin analizinde ikili kümeleme yöntemlerinin etkin kullanımını ve suç türleri ile suçların işlendiği bölgelerin eşzamanlı analizinin önemini vurgulamaktadır. Bu bulgular, suç önleme stratejilerinin geliştirilmesi ve güvenlik önlemlerinin etkin şekilde planlanmasına katkılar sağlayacaktır.

Bilim Kodu : 20513

Anahtar Kelimeler : Suç analizi, ikili kümeleme yöntemi, CC algoritması, bimax algoritması, xMotif algoritması

Sayfa Adedi : 65

Danışman : Prof. Dr. H. Hasan ÖRKÇÜ

ANALYSIS OF CRIME DATA USING BICLUSTERING METHODS

(M. Sc. Thesis)

Demet ALBASAR

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

June 2025

ABSTRACT

Crime analysis is critically important for security agencies in terms of crime predictability and taking preventive measures. With the increase in crime rates, crime data has become large and for complex datasets, making data mining techniques prominent for efficient analysis. Data mining is a method that helps discover hidden meaningful patterns and relationships from large datasets and is effectively used in crime data analysis. Many studies have been conducted on crime analysis to examine crime and the factors that cause crime together, identify their relationships, and make inferences. Through these analyses, crime types can be accurately classified, crime trends can be identified, and security measures can be planned more effectively. In particular, data mining algorithms play a significant role in understanding the relationships between crime types and the regions where these crimes are committed, especially for problems such as determining crime zones. Traditional clustering methods group data only based on rows or columns, whereas biclustering methods simultaneously group both rows and columns, allowing for more comprehensive analyses. This method is particularly effective for high-dimensional datasets. So far, many studies have been conducted in this field to examine crime and the factors that cause crime together, identify their relationships, and make inferences. In this study, the aim is to determine crime zones using biclustering methods in crime data analysis. The analysis was conducted using the CC, xMotif and Bimax algorithms on a dataset obtained from the Ministry of Justice's General Directorate of Criminal Records and Statistics, utilizing the R Project software. The obtained results highlight the effective use of biclustering methods in crime data analysis and emphasize the importance of simultaneous analysis of crime types and the regions where crimes are committed. These findings will significantly contribute to the development of crime prevention strategies and the effective planning of security measures.

Science Code : 20513

Key Words : Crime analysis, biclustering, CC algorithm, bimax algorithm, xMotif algorithm

Page Number : 65

Supervisor : Prof. Dr. H. Hasan ÖRKCÜ

TEŞEKKÜR

Yüksek lisans tezim süresince değerli rehberliğini ve engin bilgi birikimini benimle paylaşan danışmanım Prof. Dr. H. Hasan ÖRKÜ'ye en derin şükranlarımı sunarım.

Tez çalışmam süresince hukuki bilgi ve tecrübeleriyle her zaman yanımda olan, desteklerini hiçbir zaman esirgemeyen çok değerli Strateji Geliştirme Başkan Yardımcısı Kubilay İNAN, Daire Başkanı Cemil SARI ve Tetkik Hâkimi Fırat ÖLMEZ'e; ayrıca, beni daima motive eden başta Zafer KARAHAN olmak üzere kıymetli çalışma arkadaşlarıma en içten teşekkürlerimi sunarım. Ayrıca, verilerin temininde büyük katkı sağlayan Adli Sicil ve İstatistik Genel Müdürlüğü'ne de teşekkür ederim.

Her koşulda yanımda olan, moral kaynağım aileme, canım teyzem Nihal ÖZYÜKSEL'e ve dostlarıma sonsuz teşekkürlerimi iletirim.

Son olarak, hayatımdaki en değerli varlığı, sevgili kardeşim Samet ALBASAR'a buradan sonsuz teşekkürlerimi ve özlemimi iletiyorum. Onun desteği, inancı ve sevgisi her zaman yanımda oldu; yolumu aydınlattı. Fiziksel olarak yanımda olmasa da ruhu ve hatırası bana güç vermeye devam ediyor. Bu tez çalışmamı ona olan minnet ve sevgiyle tamamladım.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
ÇİZELGELERİN LİSTESİ.....	viii
ŞEKİLLERİN LİSTESİ	ix
HARİTALARIN LİSTESİ.....	xv
SİMGELER VE KISALTMALAR.....	xvi
1. GİRİŞ.....	1
2. İKİLİ KÜMELEME YÖNTEMLERİ.....	9
2.1. CC Algoritması (Cheng-Church Algoritması)	12
2.2. Bimax Algoritması (Binary Inclusion- Maximal Algorithm)	16
2.3. Plaid Algoritması.....	17
2.4. xMotif Algoritması.....	19
2.5. Quest Algoritması	20
3. UYGULAMA.....	23
3.1. Veri Yapısı ve Değişkenler	23
3.2. Tanımlayıcı İstatistikler ve İkili Kümeler	25
3.3. CC Algoritması Uygulaması	36
3.4. xMotif Algoritması Uygulaması	43
3.5. Bimax Algoritması Uygulaması.....	49
4. SONUÇ ve ÖNERİLER.....	59
ÖZGEÇMİŞ	65

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Gen açıklama verisi.....	15
Çizelge 3.1. Suç türlerine ilişkin kodlamalar.....	23
Çizelge 3.2. CC ve xMotif algoritmalarına ait suç türlerine göre illerin veri matrisi....	25
Çizelge 3.3. Bimax algoritmasına ait suç türlerine göre uyrukların veri matrisi.....	25
Çizelge 3.4. CC algoritması ikili kümelerin boyutları.....	36
Çizelge 3.5. xMotif algoritması ikili kümelerin boyutları.....	43
Çizelge 3.6. Bimax algoritması ikili kümelerin boyutları.....	54

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Sabit ikili küme gösterimi.....	11
Şekil 2.2. Toplamsal ve çarpımsal ikili küme gösterimi.....	12
Şekil 3.1. TCK göre kısımlar ve bölümlere ait hiyerarşi	26
Şekil 3.2. TCK kapsamında kısımlara ait suç verilerinin oranları	26
Şekil 3.3. TCK kapsamında bölümlere ait suç verilerinin frekansları	27
Şekil 3.4. Türkiye geneli suç türlerine ait ilk 10 suç	27
Şekil 3.5. Suç kaydı bulunan şahıslara ait medeni durum oranları	30
Şekil 3.6. Suç kaydı bulunan bireylere ait yaş gruplarına ilişkin frekanslar.....	32
Şekil 3.7. Suç kaydı bulunan bireylere ait ilk 10 meslek gruplarına ilişkin frekanslar ..	34
Şekil 3.8. CC algoritmasına ait <i>R</i> kodu.....	36
Şekil 3.9. CC algoritmasına ait ısı haritası.....	37
Şekil 3.10. Veri matrisine ait ısı haritasının genel görünümü.....	37
Şekil 3.11. CC algoritması ikili küme 1'e ait ısı haritası	38
Şekil 3.12. CC algoritması ikili küme 2'ye ait ısı haritası	39
Şekil 3.13. CC algoritması ikili küme 3'e ait ısı haritası	40
Şekil 3.14. CC algoritması ikili küme 4'e ait ısı haritası	41
Şekil 3.15. CC algoritması ikili küme 5'e ait ısı grafiği	42
Şekil 3.16. xMotif algoritması için kullanılan <i>R</i> program kodları.....	43
Şekil 3.17. xMotif algoritmasına ait ısı haritası	44
Şekil 3.18. xMotif algoritması ikili küme 1'e ait ısı haritası	45
Şekil 3.19. xMotif algoritması ikili küme 2'ye ait ısı haritası	46
Şekil 3.20. xMotif algoritması ikili küme 3'e ait ısı haritası	46
Şekil 3.21. xMotif algoritması ikili küme 4'e ait ısı haritası	47

Şekil	Sayfa
Şekil 3.22. xMotif algoritması ikili küme 5'e ait ısı haritası	48
Şekil 3.23. Suç kaydı yüksek olan ilk 10 uyruğun dağılım oranları	49
Şekil 3.24. Uyruklara ait en çok işlenen ilk 10 suç türü	51
Şekil 3.25. Bimax algoritmasına ait <i>R</i> program kodu	53
Şekil 3.26. Bimax algoritmasına ait ısı grafiği	54
Şekil 3.27. Bimax algoritması ikili küme 1'e ait ısı grafiği	55
Şekil 3.28. Bimax algoritması ikili küme 2'ye ait ısı grafiği	55
Şekil 3.29. Bimax algoritması ikili küme 3'e ait ısı grafiği	56
Şekil 3.30. Bimax algoritması ikili küme 4'e ait ısı grafiği	56
Şekil 3.31. Bimax algoritması ikili küme 5'e ait ısı grafiği	57

HARİTALARIN LİSTESİ

Harita	Sayfa
Harita 1.1. Suç kayıtlarına göre ilk 10 ve son 10 sıralarda yer alan iller	28
Harita 1.2. Suç kayıtlarına göre illerin yoğunlukları	29



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

μ	Ortalama
σ	Standart sapma
σ^2	Varyans
δ	Maksimum kabul edilebilir MSR ölçüsü
α	Ölçek faktörü

Kısaltmalar

CC	Cheng-Church algoritması
CHAID	Ki-kare otomatik etkileşim belirleme
CİMER	Cumhurbaşkanlığı iletişim merkezi
LSS	En küçük kareler yöntemi
MSR	Ortalama karesel artık
ND	Sütun örnek sayısı
NS	Satır örnek sayısı
RECAP	Bölgesel suç analizi programı
SD	Sütun özellik sayısı
SOM	Kendi kendine örgütlenen harita
TCK	Türk Ceza Kanunu
QUEST	Optimize edilmiş tümevarım arama tekniği

1. GİRİŞ

Suç, toplumun hukuki kurallarına, etik değerlerine ve normlarına aykırı davranışlar olarak tanımlanıp; demografik, ekonomik, psikolojik ve siyasal faktörlerin bir araya gelmesiyle şekillenmektedir. Suç ve suçu meydana getiren unsurlar aralarındaki ilişkilerin tespiti amacıyla suç analizi büyük önem taşımaktadır. Bu analizler sayesinde suç modelleri ve eğilimleri belirlenebilir, suç türleri sınıflandırılabilir, kaynakların daha verimli kullanılması sağlanabilir ve öngörücü analizler gerçekleştirilebilir.

Son yıllarda suç oranlarındaki artışla birlikte, suç verileri büyük hacimli veri kümeleri haline gelmiştir. Bu durum, veri analizi ve veri madenciliğinin önemini daha da artırmıştır. Veri madenciliği, büyük veri setlerinden gizli ve anlamlı desenler ile ilişkiler çıkarma süreci olarak tanımlanır ve bu sayede veriler daha etkili bir şekilde analiz edilebilir. Birçok alanda olduğu gibi suç verilerinin analizinde de veri madenciliği kritik bir rol oynamaktadır. Bu süreç, suçun zaman, mekân, neden ve türleri ile suçluların davranışlarını daha iyi anlamayı mümkün kılar. Elde edilen bilgiler, suçun önlenmesine yönelik stratejilerin geliştirilmesine katkı sağlar.

Veri madenciliği tekniklerinden biri olan kümeleme yöntemi, benzer özelliklere sahip veri noktalarını gruplandırarak analiz yapmaya olanak tanır. Geleneksel kümeleme yöntemleri genellikle yalnızca satırları veya sütunları dikkate alarak gruplama yapmaktadır. Daha kapsamlı ve detaylı analizler elde edebilmek için, özellikle gen ifade verilerinin incelenmesinde kullanılan ikili kümeleme yöntemi tercih edilmektedir. Bu yöntem, veri matrisindeki hem satırların (gözlemler) hem de sütunların (değişkenler) eş zamanlı olarak kümelenebilmesine olanak tanır.

İkili kümeleme tekniği, her kümenin farklı bir özellik alt kümesini temel alabileceği bir dizi küme oluşturmayı hedefler. Özellikle yüksek boyutlu veri setlerinde, geleneksel kümeleme yöntemlerine kıyasla daha anlamlı ve işlevsel sonuçlar sunmaktadır. Bu nedenle, suç analizi ve suç verilerinin değerlendirilmesinde ikili kümeleme yönteminin kullanımı oldukça faydalıdır.

İkili kümeleme yöntemi, suç verilerinin daha derinlemesine incelenmesini sağlayarak suç tipleri ve suçlular arasındaki gizli örüntüleri ortaya çıkarmaya yardımcı olur. Bu yöntem,

suç olayları ile suç işleyen bireyler arasındaki karmaşık ilişkileri analiz etmek için kullanılabilir. Örneğin, belirli suç türlerinin belirli demografik gruplar veya mekânsal bölgelerle nasıl kümelendiği anlaşılabilir. Ayrıca, suçların tekrar eden örüntüler oluşturup oluşturmadığı tespit edilebilir.

Suç analizi bağlamında ikili kümeleme, şu gibi alanlarda uygulanabilir:

Suç türü ve bölge analizi: Belirli suç türlerinin coğrafi konumlara göre nasıl dağıldığını incelemek için kullanılabilir. Örneğin, hırsızlık, dolandırıcılık veya şiddet suçlarının belirli mahallelerde kümelenip kümelenmediğini belirlemek mümkündür.

Suçlu profillemeye: Suç işleyen bireylerin belirli suç tipleriyle nasıl ilişkili olduğu analiz edilebilir. Özellikle tekrar suç işleme eğiliminde olan bireylerin tespiti için kullanılabilir.

Zamansal ve mekânsal örüntüler: Suç olaylarının günün hangi saatlerinde veya haftanın hangi günlerinde yoğunlaştığını belirlemek mümkündür. Bu analiz, suç önleme ve polis devriyelerinin daha verimli planlanmasını sağlar.

Sosyo-ekonomik faktörlerle ilişki: Suçların gelir seviyesi, işsizlik oranı veya eğitim düzeyi gibi değişkenlerle nasıl ilişkilendirildiği analiz edilebilir.

Bu bağlamda, ikili kümeleme yöntemi suç analizinde büyük veri setleri içerisindeki anlamlı yapıları ortaya çıkarmada güçlü bir araç olarak öne çıkmaktadır. Özellikle yapay zekâ ve makine öğrenmesi ile desteklendiğinde, suç tahmini ve önleme stratejilerinin geliştirilmesine önemli katkılar sağlayabilir. Dünyada ve Türkiye’de bu alanlarda birçok çalışma bulunmaktadır.

Brown (1998), bölgesel suç analizi programı olan ReCAP (Regional Crime Analysis Programme-Bölgesel Suç Analizi Programı) önermiştir. Bu program genellikle bir bölgedeki suçları analiz etmek, suç modellerini belirlemek ve suçla mücadele stratejilerini geliştirmek için oluşturulan bir programdır.

Adderly ve Musgrove (2001) SOM (Self Organizing Map) algoritmasını uygulayarak cinsel saldırganların suçlarını tahmin eden haritaları elde etmeye çalışmıştır. Self-Organizing Map ya da Kendi Kendine Örgütlenen Harita, Finlandiyalı bilim insanı Teuvo Kohonen

tarafından geliştirilen bir yapay sinir ağı türüdür. Bu yapay sinir ağı, benzer niteliklere sahip veri noktalarını benzer bölgelere haritalayan bir öğrenme süreciyle çalışır. SOM denetimsiz öğrenme yöntemlerinden biri olarak kullanılır ve genellikle veri madenciliği, desen tanıma, sınıflandırma, veri görselleştirme ve boyut azaltma gibi alanlarda uygulama bulur.

Chen ve diğerleri (2003), COPYLINK projesi ile çalışmışlardır. Bu proje, polis birimlerinin veri analizi ve suç soruşturmaları için kullanılan bir yazılım ve teknoloji platformudur. COPLINK, polis teşkilatlarına, farklı veri tabanları arasında bilgi paylaşımı yapma, suçluların ve suç olaylarının analizini gerçekleştirme, suç örüntülerini belirleme ve suçla mücadelede daha etkili olma imkânı sunar. Bu sistem, polis departmanlarının kullanabileceği geniş veri tabanlarına erişim sağlar ve bu veriler arasında ilişki kurarak suç olaylarını çözme, şüphelileri belirleme ve suçla mücadelede daha hızlı hareket etme olanağı tanır.

Oğuzlar (2005) tarafından Uludağ Üniversitesi ile Bursa Emniyet Müdürlüğü iş birliğinde yürütülen bir proje kapsamında, Bursa iline ait suçlu veri tabanı kullanılarak ağır suçlara karışmış bireylere ilişkin analizler gerçekleştirilmiştir. Bu çalışmada, söz konusu bireylerin benzer özelliklere göre gruplanması amacıyla SOM algoritması kullanılmış; elde edilen kümelerin yapısal özelliklerini yorumlamak üzere ise karar ağacı algoritmasından yararlanılmıştır. Analiz sonucunda ağır suç işleyen bireyler beş farklı kümeye ayrılmış ve her bir kümenin sosyo-ekonomik profili detaylı biçimde tanımlanmıştır.

Bruin, Cocx, Kosters, Laros ve Kok (2006), suçlular ve suç tanımlarını ikili karşılaştırmalar yaparak bu yöntemle suç görsel kümeleri oluşturmuş ve suçluların sınıflarının belirlenmesini sağlamıştır.

Nath (2006), veri madenciliğinde kümeleme modelini kullanmıştır. Makine öğrenmesi algoritması olan k-ortalamlar algoritması ile mekânsal haritalar yardımıyla suç örüntüsünün tespitini yapmıştır.

Sevüktekin, Oğuzlar, Aydın ve Nargeleçekenler (2007), suçluların farklı değişkenlere bağlı olarak betimlenmesi amaçlamış, ağır suç işleyen bireylere ilişkin verileri kullanmışlardır. Analizde karar ağacı algoritmalarından biri olan CHAID (Chi-square Automatic Interaction Detection) algoritmasından faydalanmışlardır. CHAID istatistiksel bir analiz tekniği ve ağaç

tabanlı bir sınıflandırma yöntemidir. Bu yöntem, bağımlı bir değişkenin kategorik olduğu durumlarda kullanılır ve bağımsız değişkenlerin bu hedef değişken üzerindeki etkilerini belirlemek için kullanılır. CHAID, veri setindeki ilişkileri ve etkileşimleri belirlemek amacıyla ağaç yapısı oluşturur. Bu ağaç, veri setindeki değişkenlerin kategorilerine göre hedef değişkeni nasıl etkilediğini gösterir. CHAID, belirli bir bağımsız değişkenin en iyi bölünmeyi yaparak hedef değişkenin farklı kategorilerine yönelik homojen alt grupları oluşturur. CHAID analizi sonucunda oluşturulan ağaç yapısı, veri setindeki değişkenler arasındaki ilişkileri görsel olarak temsil eder.

Tüzüntürk (2009), suç istatistiklerini çok değişkenli istatistiksel analiz yöntemlerinden biri olan Çok Boyutlu Ölçekleme yöntemini kullanmışlardır. Bu analizle Türkiye’de suç oranlarının düşürülmesi amaçlanmıştır. Çok boyutlu ölçekleme, veri setindeki nesnelere (örneğin, ürünler, tüketici tercihleri veya benzerlik matrisleri) göreceli konumlarını, benzerliklerini veya uzaklıklarını koruyarak düşük boyutlu bir uzayda görselleştirmek için kullanılan istatistiksel bir analiz tekniğidir.

Cömertler ve Kar (2010) tarafından yürütülen çalışmada, Türkiye'deki 81 ilin 2000 yılına ait sosyo-ekonomik göstergeleri yatay kesit veri analizi yöntemiyle incelenmiş ve bu değişkenlerin suç oranları üzerindeki etkileri araştırılmıştır.

Ma, Chen ve Huang (2010), büyük veri kümelerinden benzer olayların alt kümelerini bulmak için iki adımlı bir algoritma önermiş, belirli bir vaka kategorisine ait davranışsal niteliklerin ağırlıklarının hesaplandığı nitelik ağırlıklandırma aşaması ve niteliklerin ağırlıklarını dikkate alan AK-Modes algoritması kullanmıştır. Ak-Modes, veri madenciliği alanında kullanılan bir kümeleme (clustering) algoritmasıdır. Bu algoritma, özellikle nominal (kategorik) veri setleri üzerinde çalışmak üzere tasarlanmıştır. Nominal veriler, kategorik veya sınıflandırılabilir özelliklerdir ve sayısal olmayan değerler içerirler (örneğin, renkler, kategoriler, etiketler). Algoritma, farklı kümeleme yöntemleri kullanarak benzerlik ölçüsüne dayanarak veri noktalarını gruplara ayırır. Nominal veri türleri için özel olarak tasarlandığı için, bu tür veri setlerindeki farklı kategoriler arasındaki benzerlikleri ve farklılıkları dikkate alarak kümeleme yapabilir.

Polat, Eren ve Erbakıcı (2013) Türkiye’de ilçe düzeyinde jandarma sorumluluk bölgesindeki hırsızlık analizi için regresyon analizi ve hareketli ortalamalar tekniklerinden yararlanmışlardır.

Sea, Kim ve Youngs (2016), Güney Kore’de 111 cinsel saldırı davasında elde edilen verilere hiyerarşik kümeleme analizini uygulamıştır. Hiyerarşik kümeleme analizi, veri noktalarını benzerliklerine dayanarak gruplara ayırır ve bu grupları hiyerarşik bir yapı içinde görselleştirir.

Giray (2016), bir cezaevinin hükümlü kayıtları üzerinden gerçekleştirdiği çalışmada, benzer özellikler taşıyan hükümlü gruplarını belirlemek amacıyla iki aşamalı kümeleme analizi kullanmıştır. Araştırmada, son on yıl içinde ilgili cezaevinden tahliye edilen veya başka cezaevlerine nakledilen 1.556 hükümlüye ait demografik ve adli veriler analiz edilmiştir. İncelenen değişkenler arasında cinsiyet, yaş, suç türü ve suçun işlendiği bölge (polis veya jandarma yetki alanı) yer almıştır.

Orakcı, Ciylan, Kök ve Sevri (2016), suç kavramını ve kriminolojinin temel inceleme alanlarını esas alarak, veri madenciliği ile makine öğrenmesi yöntemlerinin suç analizine uygulanabilirliğini kapsamlı bir şekilde değerlendirmiştir. Araştırmada, literatürde yer alan çalışmalardan da yararlanılarak bu iki yöntemin birlikte kullanımının, suç analizine nasıl katkı sağladığı detaylı biçimde ele alınmıştır. Özellikle suç örüntülerinin ortaya çıkarılması, potansiyel suç bölgelerinin belirlenmesi ve olası suçların öngörülmesi gibi konularda, bu teknolojilerin emniyet birimleri açısından karar alma süreçlerine önemli düzeyde destek sağladığı ifade edilmiştir. Çalışma sonucunda, veri madenciliği ve makine öğrenmesi tekniklerinin yalnızca geçmiş verilerin incelenmesinde değil, aynı zamanda önleyici güvenlik politikalarının şekillendirilmesinde de etkili bir araç olduğu vurgulanmaktadır.

Reale, Beauregard ve Martineau (2017), Kanada’da işlenen 350 cinsel cinayet vakası örneğini kullanarak iki aşamalı kümeleme analizi ve ikili lojistik regresyonu ele almışlardır.

Stec ve Klabjan (2018), Chicago ve Portland şehirlerindeki suç verilerini kullanarak, hava durumu, nüfus sayımı ve toplu taşıma gibi ek verilerle beslenen derin sinir ağlarıyla günlük suç sayısı tahminleri yapmışlardır.

Kadar ve Pletikosa (2018), New York'taki suç verilerini Foursquare, metro ve taksi verileriyle birleştirerek, mekânsal ve zamansal özellikler kullanarak suç tahmin modelleri geliştirmiştir.

Stalidis, Semertzidis ve Daras (2018), açık polis raporları verilerini kullanarak, derin öğrenme algoritmalarının suç sınıflandırma ve tahminindeki etkinliğini incelemiştir.

Bappee, Soares Junior ve Matwin (2018), Kanada'nın Halifax şehrindeki suç verilerini kullanarak, açık sokak haritası (OSM) verileriyle mekânsal özellikler çıkararak suç tahmin modelleri geliştirmişlerdir. Bu yeni mekânsal özelliklerin, suç tahmininde önemli bir performans artışı sağladığı gözlemlenmiştir.

Ateş, Bostancı ve Msg (2020), büyük veri, veri madenciliği, makine öğrenmesi ve derin öğrenme kavramlarının suç verisi analizindeki kullanımını incelemiştir. Çalışma, suç verilerinden anlamlı bilgiler çıkarılmasının, suç öncesi, sırası ve sonrasındaki karar süreçlerinde önemli katkılar sağladığını vurgulamaktadır.

Çalışkan, Yıldız, Doğan ve Aktaş (2021), Maryland eyaletindeki suç verilerini kullanarak, Apriori ve FP-Growth algoritmalarıyla birliktelik kuralları madenciliği yapmışlardır. Elde edilen kurallar, suç türleri, zaman dilimleri ve coğrafi bölgeler arasındaki ilişkileri ortaya koymuştur.

Marin, Guerreros ve Calderon (2024), büyük veri analitiği ve kümeleme yöntemlerini kullanarak, suç türleri arasındaki ilişkileri ve eğilimleri analiz etmişlerdir. Bu yaklaşım, suçların mekânsal ve zamansal dağılımlarını daha iyi anlamaya yardımcı olmuştur.

Bu çalışmada, 2023 yılına ait Cumhuriyet Başsavcılıklarında kesinleşmiş suç verileri kullanılarak en yoğun şekilde işlenen suç türleri; iller ve uyuşuklar açısından incelenmiş, bu doğrultuda suç bölgeleri ve suç türlerine ilişkin kapsamlı bir analiz gerçekleştirilmiştir. Elde edilen bulgular, suç verilerinin değerlendirilmesinde ikili kümeleme yöntemlerinin etkinliğini ortaya koymakta ve suç türleri ile bu suçların yoğunlaştığı bölgelerin eşzamanlı analizinin önemini vurgulamaktadır. Bu sonuçların, suç önleme stratejilerinin geliştirilmesi ve güvenlik politikalarının daha etkin bir şekilde planlanmasına katkı sağlaması beklenmektedir.

Çalışmanın ikinci bölümünde, ikili kümeleme yöntemlerinin ortaya çıkış gerekçeleri ele alınmakta; CC, Bimax, Plaid, Quest ve xMotif gibi literatürde öne çıkan yöntemler ayrıntılı biçimde incelenmektedir. Farklı matematiksel ve algoritmik temellere dayanan bu yaklaşımlar aracılığıyla, veri matrislerinde anlamlı ikili kümelerin oluşturulması sağlanmaktadır.

Üçüncü bölümde ise, Adalet Bakanlığı'na bağlı Adli Sicil ve İstatistik Genel Müdürlüğü'nden temin edilen veriler kullanılarak uygulamalı analizler gerçekleştirilmiştir. Farklı veri yapılarıyla uyumlu olan CC, xMotif ve Bimax algoritmaları kapsamında yapılan analizlerde, elde edilen ikili kümeler görselleştirilerek ısı haritaları üzerinden yorumlanmıştır.

Son bölümde, elde edilen analiz sonuçları doğrultusunda kullanılan algoritmalar karşılaştırılmış; suç unsurlarını etkileyen değişkenler irdelenerek, olası sebep-sonuç ilişkileri ortaya konulmuştur. Böylece, suç olgusunun çok boyutlu yapısının anlaşılmasına katkı sunulmuş ve önleyici güvenlik politikaları için bir zemin hazırlanmıştır.



2. İKİLİ KÜMELEME YÖNTEMLERİ

Kümeleme yöntemleri, veri madenciliğinde benzer özelliklere sahip veri gruplarını belirlemek ve mevcut desenlerin özelliklerini incelemek için sıkça kullanılan önemli tekniklerdir. Bu kapsamda, hiyerarşik kümeleme ve k-ortalama gibi yöntemler, farklı küme yapılarının ortaya çıkarılmasında yaygın şekilde tercih edilmektedir. Diğer yandan, ikili kümeleme yaklaşımı, veri matrisinde hem satır hem de sütunları eş zamanlı olarak kümelendirerek daha karmaşık yapıları tanımlamaya olanak sağlar. Literatürde bu yöntem, biclustering, matris kümelemesi, co-clustering veya blok kümeleme olarak da adlandırılmaktadır.

İkili kümeleme özellikle gen ifade verilerinin analizinde ortaya çıkmıştır. Bu veriler, genlerin satırları, örneklerin ise sütunları temsil ettiği çok boyutlu matrisler şeklinde düzenlenmiştir. Böylece hem benzer gen gruplarını hem de belirli örnek kümelerini aynı anda analiz etmek mümkün olur.

Çizelge 2.1. Gen açıklama verisi

Örneklemler					
Genler	Örnek 1	Örnek 2	Örnek 3	Örnek 4	Örnek 5
Gen A	2,1	3,0	2,5	2,8	3,2
Gen B	5,5	5,7	5,8	5,9	5,6
Gen C	1,0	0,9	1,2	1,1	1,0
Gen D	3,3	3,5	3,7	3,6	3,4

Bu yapı, görselde gösterilen matris formuna karşılık gelir. Genler ve örnekler açıklama sütunları ile tanımlanabilir ve matris hücrelerindeki değerler gen ifadelerinin seviyelerini temsil eder. Bu örnekle ikili kümeleme hem genler hem de örnekler arasında ortak özellikler veya desenler keşfedilebilir.

Bu verilerde hem genleri hem de koşulları aynı anda kümeleyerek gen örüntülerini ortaya çıkartmak daha önemli hale gelmiştir. Bu amaç için ikili kümeleme algoritmaları geliştirilmiştir. Bir ikili kümeleme matrisinin yapısı aşağıdaki gibi ifade edilebilir (Kocatürk, 2018).

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdot & \cdot & \cdot & b_{1|J|} \\ b_{21} & b_{22} & \cdot & \cdot & \cdot & b_{2|J|} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ b_{|I|1} & b_{|I|2} & \cdot & \cdot & \cdot & b_{|I||J|} \end{pmatrix}$$

Burada b_{ij} i inci satır (gen) j inci sütun (koşul veya örneklem) elemanını göstermektedir. Bu veri matrisinden yola çıkarak ikili kümeleme algoritmalarında yaygın olarak kullanılan bazı formüller aşağıdaki gibidir.

$$b_{ij} = \frac{1}{|I|} \sum_{i=1}^{|I|} b_{ij} \quad (2.1)$$

$$b_{iJ} = \frac{1}{|J|} \sum_{j=1}^{|J|} b_{ij} \quad (2.2)$$

$$b_{IJ} = \frac{1}{|I||J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} b_{ij} \quad (2.3)$$

Gen verileri gösterildiği desenlere bağlı olarak, ikili kümeleme türlerinin çeşitli şekillerde tanımlanması ve sınıflandırılması yapılmıştır. Bazı türlerde ikili kümelerin değerlerini denklem ile gösterilebilmektedir. Buna göre ikili kümeleme türleri sabit değerli ikili küme, satır ya da sütunlarda sabit değerli ikili küme, toplamsal ve çarpımsal model ikili küme başlıkları ile kategorize edilebilir.

Sabit ikili küme türleri

Sabit değerli ikili kümeler sabit, satır sabit ve sütun sabit olmak üzere kendi içerisinde üçe ayrılmaktadır. Sabit ikili küme elemanları $b_{ij} = \pi$, satır sabit ikili kümeler $b_{ij} = \pi + \beta_j$ (toplamsal) veya $b_{ij} = \pi \times \beta_j$ (çarpımsal) şeklinde, sütun sabit ikili kümelerin değerleri ise $b_{ij} = \pi + \beta_i$ (toplamsal) veya $b_{ij} = \pi \times \beta_i$ (çarpımsal) şeklinde değerler almaktadır.

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

(a)

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

(b)

1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5

(c)

Şekil 2.1. Sabit ikili küme gösterimi

Burada; (a) sabit değerler, (b) sabit satırlar ve (c) sabit sütunları göstermektedir.

Toplamsal ve çarpımsal ikili küme türleri

Aguilar-Ruiz (2005), gen verileri içerisindeki ikili kümeler için sabit ikili kümelerden farklı bir yapı önermiştir. Bu yapıları ikili küme elemanları arasındaki sayısal ilişkiyi inceleyerek tanımlamıştır. Sabit olmayan her bir ikili küme (sütun elemanları aynı kalmak koşuluyla) satır elemanları için parametre değeri π_i olsun.

Toplamsal model için ikili kümelerin satır elemanlarına, her bir sütuna β_j katsayısı eklenerek matematiksel model ($b_{ij} = \pi_i + \beta_j$) oluşturulur. İyi bir toplamsal ikili küme yapısı için satırlar arasındaki sayısal ilişki benzer olmalıdır.

Çarpımsal model için de benzer durum geçerlidir. Çarpımsal modelin parametresi α_j katsayısı her bir sütun ile çarpılarak model ($b_{ij} = \pi_i \times \alpha_j$) oluşturulur.

Her iki modelin ikili küme içerisinde aynı anda bulunması durumunda ise sütun elemanlarına hem toplamsal modeldeki β_j katsayısı eklenir hem de çarpımsal modeldeki α_j katsayısı çarpılır. Bu durumda toplamsal ve çarpımsal ikili küme modeli ($b_{ij} = \pi_i \times \alpha_j + \beta_j$) oluşturulur.

Toplamsal ve çarpımsal ikili kümelerin örnek gösterimi Şekil 2.2’de gösterilmiştir.

1	2	0	3	2
2	3	1	4	3
3	4	2	5	4
4	5	3	6	5
5	6	4	7	6

(a)

1	2	6	3	9
2	4	12	6	18
3	6	18	9	27
4	8	24	12	36
5	10	30	15	45

(b)

1	3	7	5.3	15
2	5	13	7.3	21
3	7	19	9.3	27
4	9	25	11	32
5	11	31	13	38

(c)

Şekil 2.2. Toplamsal ve çarpımsal ikili küme gösterimi

Şekil 2.2’de görüldüğü üzere; (a) toplamsal, (b) çarpımsal iken (c) toplamsal ve çarpımsal ikili kümeleri göstermektedir.

Görüldüğü üzere ikili küme türlerinin farklılık göstermesi sebebiyle birçok ikili kümeleme algoritması mevcuttur. En çok kullanılan algoritmalar; Cheng ve Church (CC), Bimax, Plaid, Quest ve xMotif algoritmalarıdır.

2.1. CC Algoritması (Cheng-Church Algoritması)

Cheng ve Church (2000) tarafından geliştirilen CC algoritması, gen ekspresyon verilerinde ilk kullanılan ikili kümeleme yöntemlerinden biridir. Algoritmanın temel amacı, ortalama kare artığı (Mean Squared Residue-MSR) skorunu en aza indirerek, genler (satırlar) ve koşullar (sütunlar) arasında yüksek benzerlik gösteren alt kümeleri belirlemektir.

Burada MSR, ikili kümedeki tüm elemanların varyansını ifade etmektedir. Bu değer küçüldükçe satır ve sütun değerleri tutarlıdır ve bulunan ikili kümeleme anlamlıdır. MSR skoru aşağıda yer alan formülle hesaplanmaktadır.

$$H(I, J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} (a_{ij} - a_{iJ} - a_{iI} + a_{IJ})^2 \quad (2.4)$$

$$a_{iJ} = \frac{1}{|J|} \sum_{j \in J} a_{ij} \quad (2.5)$$

$$a_{iI} = \frac{1}{|I|} \sum_{i \in I} a_{ij} \quad (2.6)$$

$$a_{IJ} = \frac{1}{|I||J|} \sum_{i \in I, j \in J} a_{ij} = \frac{1}{|I|} \sum_{i \in I} a_{ij} = \frac{1}{|J|} \sum_{j \in J} a_{ij} \quad (2.7)$$

Yukarıdaki eşitliklerde bulunan a_{ij} ikili kümelerin artık değeri iken, (I,J) alt matrisinde yer alan a_{ij} satır ortalaması, a_{ij} sütun ortalaması ve a_{IJ} de toplam ortalamadır. Algoritmada bir δ sabiti belirlenerek H değeri bu sabitle karşılaştırılır. $\delta \geq 0$ olmak üzere $H(I,J) \leq \delta$ ise elde edilen ikili küme δ -ikili küme olarak ifade edilmektedir. $H(I,J)=0$ ise bulunan ikili küme sabit bir ikili kümedir.

Algoritmanın ilk adımında veri matrisi için H değeri hesaplanır ve çoklu düğüm silme işlemi için α eşik değeri belirlenir. $H > \alpha$ ile satır ve sütun silme işlemi yapılmaktadır. H değerini azaltmak amacıyla ikinci adımda da satır ya da sütunlar tek tek silinmeye devam eder. Satır sütun silme işlemlerinden sonra ikili kümenin boyutunu büyütmek amacıyla H değeri δ değerini aşmayacak şekilde rastgele satır ya da sütun eklenir. Bu adım H değeri δ değerine yaklaşıncaya kadar devam eder. İkili kümenin boyutu sabit kaldığında algoritma sonlanmaktadır. Veri matrisi güncellenerek ikinci iterasyona geçilir.

Tekli düğüm silme algoritması

Bu algoritma, düşük benzerlik gösteren genleri (satırları) veya koşulları (sütunları) ikili kümeden çıkarmak için kullanılır.

Algoritmanın girdileri, genler (satırlar) ve koşullar (sütunlar) arasındaki ilişkiyi temsil eden gerçek sayılardan oluşan A matrisi ile maksimum kabul edilebilir ortalama kare artık skoru $\delta \geq 0$ parametresidir.

Eğer ikili kümelerin MSR değeri δ 'dan büyükse, kümeler arasındaki farklılık fazla olduğu için düğüm silme işlemi başlatılır.

Algoritmanın çıktısı; A_{IJ} matrisinin belirli satır ve sütunları silindikten sonra kalan alt matrisidir. δ 'dan büyük olmayan satır kümesi I ve sütun kümesi J olan alt matrisinin δ -ikili kümesidir.

I ve J, orijinal verideki gen (satır) ve koşul (sütun) kümeleri olarak başlatılır. A_{IJ} , başlangıçta A matrisinin tamamıdır. Ancak, bu matris düğümler (satırlar/sütunlar) silindikçe küçülür.

Algoritmanın adımları (tekrar etme döngüsü) ise şu şekildedir:

1. Öncelikli olarak MSR değeri hesaplanır. Her $i \in I$ için a_{iJ} , her $j \in J$ için a_{iJ} , a_{IJ} ve $H(I,J)$ hesaplanır. Eğer $H(I,J) \leq \delta$ ise A_{IJ} 'ye dönülür. $H(I,J) > \delta$ ise en fazla hata yapan satır veya sütun silinir.
2. Satır ve sütun için hata payı (d) hesaplanır. Bu değer, satır veya sütunda ikili küme için ne kadar hata yaptığını göstermektedir. $d(i)$ değeri büyükse, o satır ya da o sütun ikili kümeye uymuyor demektir. I ve J yeni ikili kümelere göre güncellenir ve MSR değeri yeniden hesaplanır. $MSR \leq \delta$ olduğunda iterasyon tamamlanır.

Çoklu düğüm silme algoritması

Bu algoritma, bir matris içinde belirli bir hata eşliğini aşan satır ve sütunları iteratif olarak silerek bir alt matris elde etmeyi amaçlamaktadır.

Algoritmanın girdileri; A gerçekte sayılardan oluşan bir matris, δ sıfırdan büyük maksimum ortalama artık kare değeri ve $\alpha \geq 1$ çoklu düğüm silme için bir eşik değeridir.

Algoritmanın çıktıları; seçilen satır kümesi I ve sütun kümesi J olan A_{IJ} alt matrisidir. Bu alt matrisin hata değeri δ 'dan büyük olmamalıdır.

Başlangıç kümesi I ve J olan tüm satır ve sütunları içermektedir. Başlangıç matrisi $A_{IJ} = A$ 'dır.

Algoritmanın adımları ise şu şekildedir:

1. Her $i \in I$ için a_{iJ} , her $j \in J$ için a_{iJ} , a_{IJ} ve $H(I,J)$ hesaplanır. Eğer $H(I,J) \leq \delta$ ise A_{IJ} 'ye dönülür.
2. Satır için $\frac{1}{|J|} \sum_{j \in J} (a_{ij} - a_{iJ} - a_{ij} + a_{IJ})^2 > \alpha H(I,J)$ koşulu sağlanıyorsa ilgili satır silinir ve satır silinmesi durumunda ortalamalar yeniden hesaplanır. Sütun için ise $\frac{1}{|I|} \sum_{i \in I} (a_{ij} - a_{iJ} - a_{ij} + a_{IJ})^2 > \alpha H(I,J)$ koşulu sağlanıyorsa ilgili sütun silinir. Hiç satır veya sütun silinmiyorsa Algoritma 1'e dönülür.

Algoritma 3 (Düğüm ekleme)

CC algoritmasında düğüm ekleme işlemi kümelenen alt matrisin genişletilerek daha iyi bir yapıya ulaşmasını sağlamak uygulanmaktadır. Bu sayede kümelenen blokları genişletebilir ve hata azaltılabilir. Özellikle verinin gizli yapısını ortaya çıkarmak ve kümelerin kalitesini artırmak açısından önemlidir.

Algoritmanın girdileri; gerçek sayılardan oluşan bir A matrisi ve başlangıçta belirlenmiş I ve J adında bir δ - ikili kümedir.

Algoritmanın çıktıları ise başlangıç kümelerinin alt veya eşit oluşturacak şekilde I' ve J' 'dir. Burada amaç $H(I', J') \leq H(I, J)$ olacak şekilde kümeleri genişletmektir.

Algoritmanın adımları:

1. Her $i \in I$ için a_{ij} , her $j \in J$ için a_{Ij} ve $H(I, J)$ hesaplanır.
2. $\frac{1}{|I|} \sum_{i \in I} (a_{ij} - a_{iJ} - a_{Ij} + a_{IJ})^2 \leq \alpha H(I, J)$ koşulunu sağlayan $j \notin J$ sütunu eklenir.
3. a_{ij} , a_{Ij} ve $H(I, J)$ değerleri yeniden hesaplanır.
4. $\frac{1}{|J|} \sum_{j \in J} (a_{ij} - a_{iJ} - a_{Ij} + a_{IJ})^2 \leq \alpha H(I, J)$ koşulunu sağlayan $i \in I$ satırı eklenir.
5. Her bir satır için $i \in I$ 'da değilse; $\frac{1}{|J|} \sum_{j \in J} (-a_{ij} + a_{iJ} - a_{Ij} + a_{IJ})^2 \leq \alpha H(I, J)$ işlemi eklenir.
6. Eğer yeni bir satır veya sütun eklenmediyse, I ve J kümelerini I' ve J' hâlleriyle döndürülür.

Algoritma 4 (Belirli bir sayıda ikili küme bulma)

Bu algoritma ikili kümeleme yönteminin bir türevi olup, önceden belirlenen sayıda n tane δ - ikili küme elde etmek için uygulanır.

Algoritmanın girdileri; eksik elemanlara sahip gerçek sayılardan oluşan bir A matrisi, çoklu düğüm silme için eşik değer olan α , maksimum kabul edilebilir ortalama karesel artış değeri $\delta \geq 0$ ve bulunacak δ -ikili kümelerinin sayısı n'dir.

Algoritmanın çıktıları ise; n tane δ - ikili kümedir.

Algoritmanın adımları:

1. A matrisindeki eksik değerler, mevcut verilerin aralığında yer alan rastgele sayılarla değiştirilerek başlanır. Burada A' matrisi A matrisinin kopyası olarak oluşturulur.
2. A' matrisine δ ve α eşik değerleri kullanılarak Algoritma 2 uygulanır. Eğer bir satır veya sütunun boyutu çok küçükse, Algoritma 2 uygulanmaz. Silinmeyen katsayılardan oluşan yeni matris B elde edilir.
3. Yeni bir küme elde etmek için, B ve δ üzerine Algoritma 1 uygulanır. Elde edilen yeni matris C olarak adlandırılır.
4. A' ve C matrislerine Algoritma 3 uygulanarak elde edilen yeni matris D olarak adlandırılır.
5. D matrisi analiz edilerek, D'deki rastgele sayılar A'daki öğelerle değiştirilir.

Cheng-Church algoritmasının bazı önemli eksiklikleri bulunmaktadır. İlk olarak, elde edilen sonuçlara herhangi bir istatistiksel anlamlılık değeri atanmamaktadır. δ değeri sabit olduğundan, başlangıç matrisi yeterince büyük seçildiğinde, koşulu sağlayan rastgele boyutlarda ikili kümeler bulunma olasılığı oldukça yüksektir. İkinci olarak, algoritma her iterasyonda yalnızca tek bir ikili küme tespit ettiği için veriye gereksiz rastgele sayılar eklenebilir ve bu da anlamsız kümelerin oluşmasına yol açabilir. Ayrıca, algoritma eksik veri içeren veri setleriyle doğrudan çalışamaz. Eğer veri setinde eksik değerler varsa, bu değerlerin rastgele sayılarla değiştirilmesi gerekmektedir.

2.2. Bimax Algoritması (Binary Inclusion- Maximal Algorithm)

Bimax algoritması, yalnızca 0 ve 1 değerlerinden oluşan matrislerde, 1'lerden meydana gelen alt matrisleri belirlemeye yönelik olarak Prelic ve arkadaşları (2006) tarafından geliştirilmiştir. Büyük ölçekli ikili veri kümelerinde yüksek hızda çalışabilen bu algoritma, diğer yöntemlerden farklı olarak aykırı değerlerden etkilenmez. Bimax, $m \times n$ boyutundaki bir matris M'de yer alan tüm olası ikili kümeleri belirlemek için özyinelemeli bir böl ve yönet stratejisini benimser.

Algoritma, 0 ve 1'lerden oluşan bir veri matrisinden rastgele bir satır seçilerek başlatılır. Daha sonra, sütun kümesi, 0 ve 1 değerlerine göre iki alt kümeye ayrılır. Aynı bölme işlemi satırlar için de uygulanır. Her adımda, satır ve sütunların sıralamaları yeniden düzenlenerek alt matrisler oluşturulur. Bu süreç, en büyük alt matris elde edilene kadar devam eder. Bölme işlemi tamamlandıktan sonra, 0 içeren satır ve sütunlar kaldırılarak yalnızca 1'lerden oluşan bir ikili küme oluşturulur.

Algoritmanın girdisi $M \times N$ boyutlu $E_{M \times N}$ veri matrisidir.

Bimax algoritmasının başlangıç adımı için E matrisinden; U ve V alt matrisleri elde etmek için rastgele bir m satır elemanı seçilir.

Belirlenen kurallara göre rastgele veya belirli bir yöntemle matrisin sütunları N_U ve N_V olmak üzere iki alt kümeye ayrılır. Sadece N_U alt kümesinde bulunan sütunlarla uyumlu satırları içeren M_U alt kümesi oluşturulur. Hem N_U hem de N_V alt kümelerinin kesişiminde yer alan satırlar M_W kümesine alınır. Sadece N_V kümesindeki sütunlarla uyumlu olan satırlar M_V kümesine atanır.

Satırlardan elde edilen M_U , M_W ve M_V kümelerine göre, sütun elemanları da tekrar gözden geçirilerek N_U ve N_V kümeleri güncellenir.

U ve V alt matrisleri birbirinden tamamen ayrışana kadar bu bölme işlemi tekrar edilir. Böylece matris farklı gruplara bölünerek veri kümelerinde anlamlı ilişkiler ortaya çıkarmış olur.

2.3. Plaid Algoritması

Plaid modeli, toplamsal (additive) bir ikili kümeleme tekniği olup, veri matrisindeki elemanları farklı katmanlar olarak ele alır. Lazzeroni ve Owen (2000) tarafından geliştirilen bu algoritma, satır ve sütun etkilerini dikkate alarak ikili kümeleri belirlemeye yönelik doğrusal bir model oluşturur. Algoritmanın temel mantığı, veri matrisini ikili kümelere karşılık gelen katmanların doğrusal bir fonksiyonu şeklinde ifade etmektir:

$$a_{ij} = \sum_{k=0}^K \theta_{ijk} \rho_{ik} K_{jk} \quad (2.8)$$

Burada; K ikili küme sayısını ifade eder. θ_{ijk} ikili küme k 'nin matrisin i . satır ve j . sütunu üzerindeki katkısını gösterir. ρ_{ik} ve K_{jk} ise sırasıyla satır ve sütunun belirli bir ikili kümeye ait olup olmadığını gösteren ikili (binary) değişkenlerdir.

Plaid modeli, diğer ikili kümeleme algoritmalarından farklı olarak ikili kümelerin örtüşmesine izin vermektedir. Bu sayede veri matrisindeki her hücre birden fazla ikili kümeye ait olabilir.

Modelde θ_k değişkeninin farklı kullanımlarıyla farklı türde ikili kümeler tanımlanabilir. $\theta_{ijk} = \mu_k$ ise model K adet sabit ikili kümeleri tanımlar. $\theta_{ijk} = \mu_k + \alpha_{ik}$ ise sabit satırlara sahip; $\theta_{ijk} = \mu_k + \beta_{jk}$ olduğunda sabit sütunlara sahip ikili kümelerdir. Son olarak satır ve sütun etkilerini içeren toplamsal model ise $\theta_{ijk} = \mu_k + \alpha_{ik} + \beta_{jk}$ şeklinde tanımlanır.

Algoritmada, satır ve sütunlar bir araya getirilerek büyük veri matrisi içinde dikdörtgenel bir alan oluşturulur ve bu alan bir alt matris şeklinde kümeleme (blok) işlemine tabi tutulur. Her blok, içindeki değerlerin birbirine yakın olmasıyla karakterize edilir ve grafik üzerinde benzer renk tonlarıyla temsil edilir. Cebirsel olarak, model şu şekilde ifade edilir.

$$Y_{ij} = \mu_0 + \sum_{k=1}^K \mu_k r_{ik} c_{jk} \quad (2.9)$$

Eşitlik 2.2'deki μ_k değeri, ikili küme için veri setindeki tüm satır ve sütun elemanlarını içeren durumlarda kullanılan doğrusal model denklemdir. Gen ifade veri matrislerinde, belirli gen kümeleri bazı örnekler üzerinde etkili olurken, örnekler kümesi içinde de bazı genler belirleyici rol oynar. Bu etkileşimi daha iyi ifade edebilmek için doğrusal denkleme belirli parametrelerin eklenmesi gerekmektedir.

$$Y_{ij} = \mu_0 + \sum_{k=1}^K (\mu_k + a_{ik}) r_{ik} c_{jk} \quad (2.10)$$

$$Y_{ij} = \mu_0 + \sum_{k=1}^K (\mu_k + b_{jk}) r_{ik} c_{jk} \quad (2.11)$$

$$Y_{ij} = \mu_0 + \sum_{k=1}^K (\mu_k + a_{ik} + b_{jk}) r_{ik} c_{jk} \quad (2.12)$$

Eşitlik 2.10'da yer alan a_{ik} parametresi satır etkisini gösterirken; eşitlik 2.11'de yer alan b_{jk} parametresi sütun etkisini göstermektedir. Eğer $|\mu_k + \alpha_{ik}|$ değeri büyükse, satır

elemanlarının etkisi daha baskın olurken, $|\mu_k + \beta_{jk}|$ değeri büyükse, sütun elemanlarının etkisi daha belirgin hale gelir. Bu modeli çözmek için Busygin ve arkadaşları (2008), sezgisel ve iteratif tabanlı algoritmalar geliştirmiştir.

Algoritma iteratif bir süreçle başlamaktadır. Modelde, A matrisinden başlıca etkileri çıkararak bir artıklar matrisi olarak tanımlanan matrisi oluşturulur. Bu adım modelde açıklanamayan kısmı temsil etmektedir.

r_{ik} (sıra göstergesi) ve c_{jk} (sütun göstergesi) için başlangıç değerleri 0 olarak atanır. Önceki iterasyonun değerleri kullanılarak en küçük kareler yöntemi ile tahmin ediciler hesaplanır. Bulunan μ^n , $a_{ik}^{(n)}$ ve $b_{ik}^{(n)}$ tahmin edicileri ile $c_{jk}^{(n)}$ sütun göstergesi tahmin edilir. Sütun göstergesi güncellendikten sonra aynı yöntemle bu defa $r_{ik}^{(n)}$ sıra göstergesi tahmin edilir. Eğer $0.5+n/2 (N-T) > 0.5$ ise yeni bir $r_{ik}^{(n)}$ veya $c_{jk}^{(n)}$ etkilerinden biri eklenebilir. Eğer $0.5+n/2 (N-T) < 0.5$ ise sıra veya sütun göstergesinden herhangi biri çıkartılmalıdır. Güncellenmiş gösterge değerleri ile μ^{N+1} , $a_{ik}^{(N+1)}$ ve $b_{ik}^{(N+1)}$ tahmin edicileri tekrar hesaplanır. Aday küme için LSS (Least Squares Sum) değerini şu formülle gösterilmektedir:

$$LSS_{aday} = \sum_{i,j} (\mu + \alpha_i + \beta_j) r_i c_j \quad (2.13)$$

Bu değer, model üzerindeki en uygun ikili kümeyi belirlemek için kullanılır.

Maksimum LSS değerine sahip sıra ve sütun çifti seçilir ve en iyi ikili küme belirlenerek algoritma tekrar başa döner.

2.4. xMotif Algoritması

Murali ve Kasif (2003) tarafından geliştirilen bu algoritma, belirli bir sütun alt kümesi içinde korunmuş genlerin alt kümesi olarak tanımlanır. Burada “korunmuş” terimi, sütunların belirli bir sütun kümesi boyunca aynı seviyede olması anlamına gelmektedir.

Algoritma, verideki tüm örnekleri (sütunları) ve sınıfları kapsayan en büyük korunmuş gen motiflerini bulmayı amaçlar ve kesikli veri kümesi üzerinde çalisan nondeterministik

açgözlü bir yöntemdir. Temel amacı, belirli deneysel koşullar (sütunlar) boyunca genlerin (satırların) ifade seviyelerinin korunmuş olduğu en büyük kümeleri belirlemektir.

Algoritmanın girdisi, $S \times D$ boyutlu sıralı bir veri matrisi olan A matrisidir. Burada S satır sayısını (genler), D ise sütun sayısını (koşullar/örnekler) ifade etmektedir.

Algoritmaya; A matrisinin satır elemanlarından rastgele bir örnek (c) seçilerek başlanır. Seçilen c satırında motif oluşturma işlemi şekillenir. A matrisinin sütun elemanlarından rastgele D alt kümesi seçilir. Bu seçilen D alt kümesi, motifin oluşturulmasında temel koşulların belirlenmesini sağlar. Bu adımlarda amaç, tüm genler arasında belirli koşullarda ortak bir desen bulmaktır.

Seçilen D alt kümesi içinde, g geninin belirlenen koşullara uyup uymadığı kontrol edilir. Eğer genişletilen motif içindeki genler, belirlenen D alt kümesi içinde korunmuşsa, yani aynı ifade seviyesinde kalıyorsa, bu genler G_{ij} kümesine eklenir. C_{ij} , G_{ij} kümesindeki tüm gen ifadelerinde c ile uyumlu örnek kümesi olarak alınır. Burada C_{ij} ; G_{ij} kümesindeki genlerin, seçilen c genine uyumlu olan örneklerden oluşan bir alt kümesidir. Bu adımda, tüm genlerin motifin merkezine (c) uyup uymadığı test edilir. Eğer genler motifle uyumlu ise, bu genler C_{ij} kümesine dâhil edilir. Eğer C_{ij} kümesi, belirlenen minimum boyuttan küçükse, bu motif geçersiz sayılır ve kümeleme işleminden çıkarılır. Algoritma $|G_{ij}|$ maksimum olana kadar devam eder.

Algoritma örtüşen kümelere izin vermektedir ve ikili kümelerin önceden belirlenmesine gerek yoktur. Ancak bu model örtüşen kümelerde iyi bir performans göstermez. Ayrıca verileri kesikli değerlere dönüştürdüğünden aykırı değerlerden etkilenir.

2.5. Quest Algoritması

OUEST, "Optimized Exhaustive Search Technique" veya benzer bir açılım taşıyan, biclustering yapmak için kullanılan bir algoritmadır. Algoritma alt kümeleri tararken tüm olasılıkları veya büyük bir alt kümesini optimize ederek araştırabilmektedir.

Algoritmada örtüşen ve örtüşmeyen ikili kümeler oluşturabilir. Optimizasyon temellidir ve kapsamlı arama gerçekleştirmektedir.

Quest algoritmasının çalışma prensibinde ilk olarak başlangıç olarak veri matrisi alınır. Burada satırlar genleri, sütunlar ise deney koşullarını temsil eder. Veri matrisinin hem satır hem sütun alt kümeleri oluşturulur. Optimizasyon kriteri değerlendirilerek (örneğin, ikili küme içindeki genlerin belirli bir korelasyon veya modelle uyumu ölçülür) kapsamlı arama veya optimize edilmiş arama stratejileriyle en uygun alt kümeler bulunur.





3. UYGULAMA

Uygulamaya ait suç verileri, Adalet Bakanlığı Adli Sicil ve İstatistik Genel Müdürlüğü'nden, kişilerin hak ve özgürlükleri korunarak, CİMER üzerinden resmi izinle temin edilmiştir. Veriler, 2023 yılı itibarıyla Cumhuriyet Başsavcılıklarında kesinleşmiş suçları içermektedir.

Analiz kapsamında Türk Ceza Kanunu (TCK-5237), Ateşli Silahlar ve Bıçaklar ile Diğer Aletler Hakkında Kanun (6136) ve Kaçakçılıkla Mücadele Kanunu (5607) olmak üzere 3 kanun türünde çalışılmıştır.

3.1. Veri Yapısı ve Değişkenler

CC algoritması ve xMotif algoritmasında kullanılan veri seti 696.285 suç kaydından oluşmaktadır. Veri seti, suç türlerine ve suçluların demografik özelliklerine (*uyruk, cinsiyet, yaş, medeni durum, eğitim durumu, meslek, il ve ilçe*) ait değişkenleri içermektedir. Suçlara ilişkin veriler 58 farklı suç türünden oluşmaktadır.

Çizelge 3.1. Suç türlerine ilişkin kodlamalar

A1_1a	Göçmen kaçakçılığı ve insan ticareti
A2_1a	Kasten öldürmede nitelikli haller
A2_1b	Kasten öldürme
A2_1c	Taksirle öldürme
A2_2a	Kasten yaralama
A2_2b	Taksirle yaralama
A2_2c	Vücut dokunulmazlığına karşı diğer suçlar
A2_3a	Cinsel saldırı
A2_3b	Cinsel taciz
A2_3c	Cinsel dokunulmazlığa karşı diğer suçlar
A2_4a	Konut dokunulmazlığın ihlali
A2_4b	Tehdit
A2_4c	Kişileri hürriyetinden yoksun kılma
A2_4d	Kişilerin huzur ve sükununu bozma
A2_4e	Hürriyete karşı diğer suçlar
A2_5a	Hakaret
A2_6a	Dolandırıcılık
A2_6b	Güveni kötüye kullanma
A2_6c	Hakkı olmayan yere tecavüz
A2_6d	Hırsızlık
A2_6e	Karşılıksız yararlanma
A2_6f	Mala zarar verme

Çizelge 3.1. (devamı) Suç türlerine ilişkin kodlamalar

A2_6g	Mala zarar vermenin nitelikli halleri
A2_6h	Nitelikli dolandırıcılık
A2_6i	Nitelikli hırsızlık
A2_6j	Nitelikli yağma
A2_6k	Malvarlığına karşı diğer suçlar
A2_7a	Kişilere karşı diğer suçlar
A3_1a	Trafik güvenliğini tehlikeye sokma
A3_1b	Genel tehlike yaratan diğer suçlar
A3_2a	İmar kirliliğine neden olma
A3_2b	Çevreye karşı diğer suçlar
A3_3a	Uyuşturucu veya uyarıcı madde imal ve ticareti
A3_3b	Kullanmak için uyuşturucu veya uyarıcı madde satın almak, kabul etmek veya bulundurmaya ya da uyuşturucu veya uyarıcı madde kullanmak
A3_3c	Kamunun sağlığına karşı diğer suçlar
A3_4a	Mühür bozma
A3_4b	Resmî belgede sahtecilik
A3_4c	Parada sahtecilik
A3_4d	Kamu güvenine karşı diğer suçlar
A3_5a	Fuhuş
A3_5b	Kumar oynanması için yer ve imkân sağlama
A3_5c	Genel ahlaka karşı diğer suçlar
A3_6a	Aile düzenine karşı suçlar
A3_6b	Bilişim alanında suçlar
A3_6c	Ekonomi, sanayi ve ticarete ilişkin suçlar
A3_6d	Kamu barışına karşı suçlar
A3_6e	Ulaşım araçlarına veya sabit platformlara karşı suçlar
A4_1a	Görevini yaptırmamak için direnme
A4_1b	Kamu idaresinin güvenilirliğine ve işleyişine karşı diğer suçlar
A4_2a	Silahlı örgüt
A4_2b	Anayasal düzene ve bu düzenin işleyişine karşı diğer suçlar
A4_3a	Hükümlü veya tutuklunun kaçması
A4_3b	İnfaz kurumuna veya tutukevine yasak eşya sokmak
A4_3c	Adliyeye karşı diğer suçlar
A4_4a	Topluma karşı diğer suçlar
B1	Mermileri satın alma, taşıma veya bulundurmaya ilişkin suçlar
B2	6136 sayılı kanuna muhalefet diğer suçlar
C1	Kaçakçılık suçları

Veri matrisine ait yapı Çizelge 3.2’de verilmiştir. Bu veri matrisinin satırları suç türlerini sütunları ise illeri göstermektedir. Matrisin elemanları i . suçun j . ilde meydana gelen suçun sayısını göstermektedir. Böylece analizde 58x81 boyutunda bir veri matrisi kullanılmıştır.

Çizelge 3.2. CC ve xMotif algoritmalarına ait suç türlerine göre illerin veri matrisi yapısı

	İller			
Suç Kodu/Türü	<i>Adana</i>	<i>Adıyaman</i>	<i>Zonguldak</i>
<i>AI_1a</i>				
<i>AI_2b</i>				
<i>AI_2b</i>				
.				
.				
.				
<i>CI</i>				

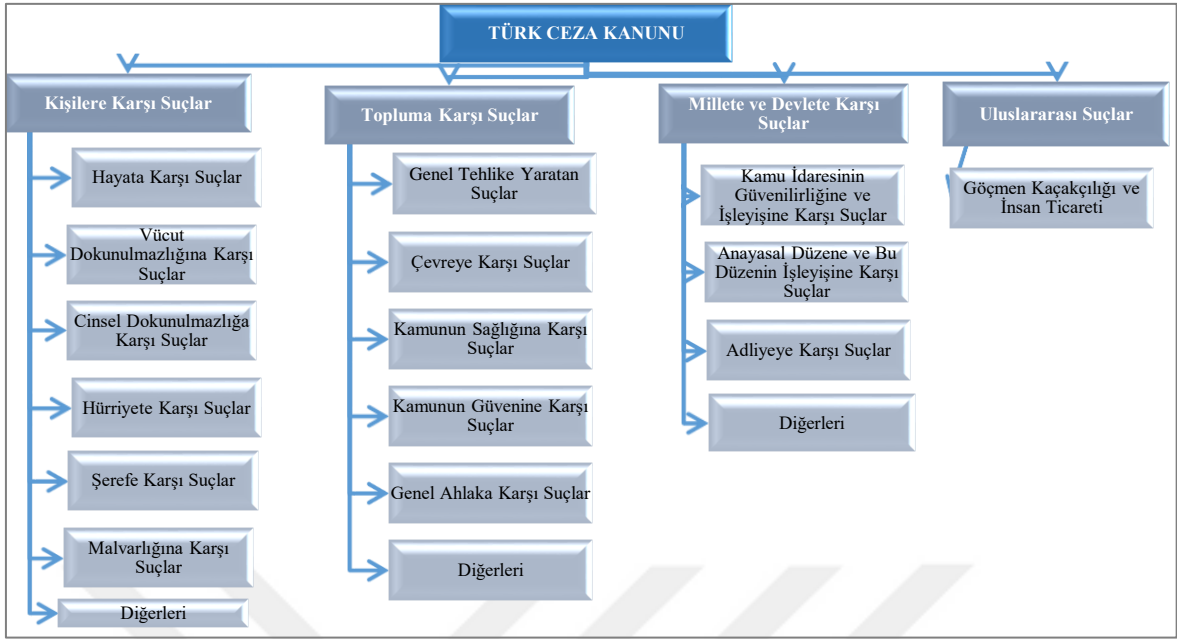
Bimax algoritması analizde ise ikili (binary) değerlerden oluşan ve satırlarda suç türlerinin, sütunlarda ise uyrukların yer aldığı 100x58 boyutlarında bir matris kullanılmıştır. Veri seti, toplam 7.429 suç kaydını içermektedir. Bu yapı üzerinden uyruk ve suç türleri bakımından detaylı analiz gerçekleştirilmiştir.

Çizelge 3.3. Bimax algoritmasına ait suç türlerine göre uyrukların veri matrisi yapısı

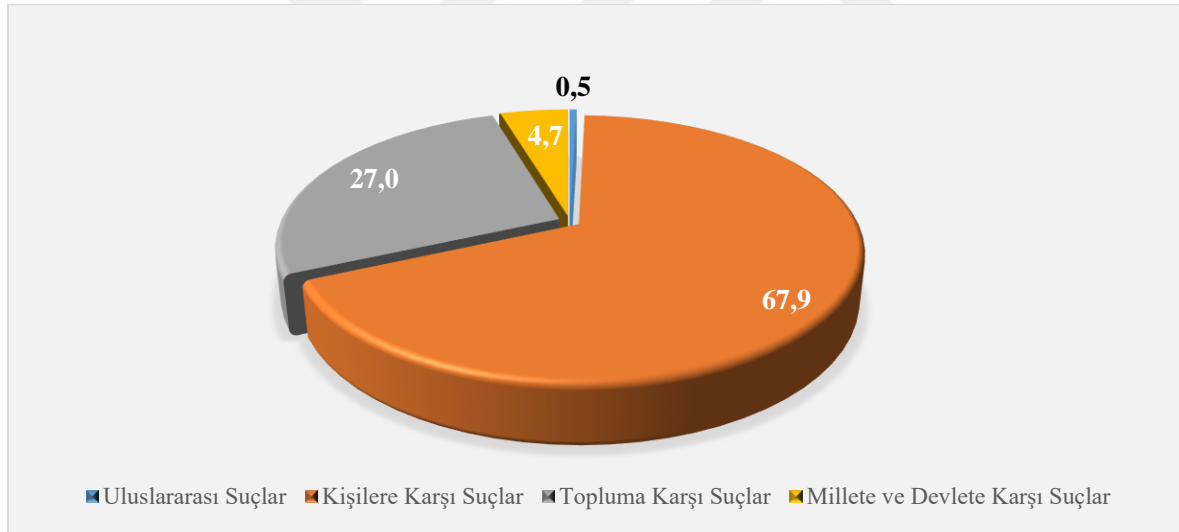
	Uyruk			
Suç Kodu/Türü	<i>Afganistan</i>	<i>Almanya</i>	<i>Zambiya</i>
<i>AI_1a</i>				
<i>AI_2b</i>				
<i>AI_2b</i>				
.				
.				
.				
<i>CI</i>				

3.2. Tanımlayıcı İstatistikler ve İkili Kümeler

Çalışma kapsamında üç farklı veri yapısını temsil eden CC (sürekli), Bimax (ikili) ve xMotif (kesikli) algoritmaları kullanılmıştır. Analizler, her algoritmanın gerektirdiği veri yapısına uygun biçimde R Project programında gerçekleştirilmiştir.

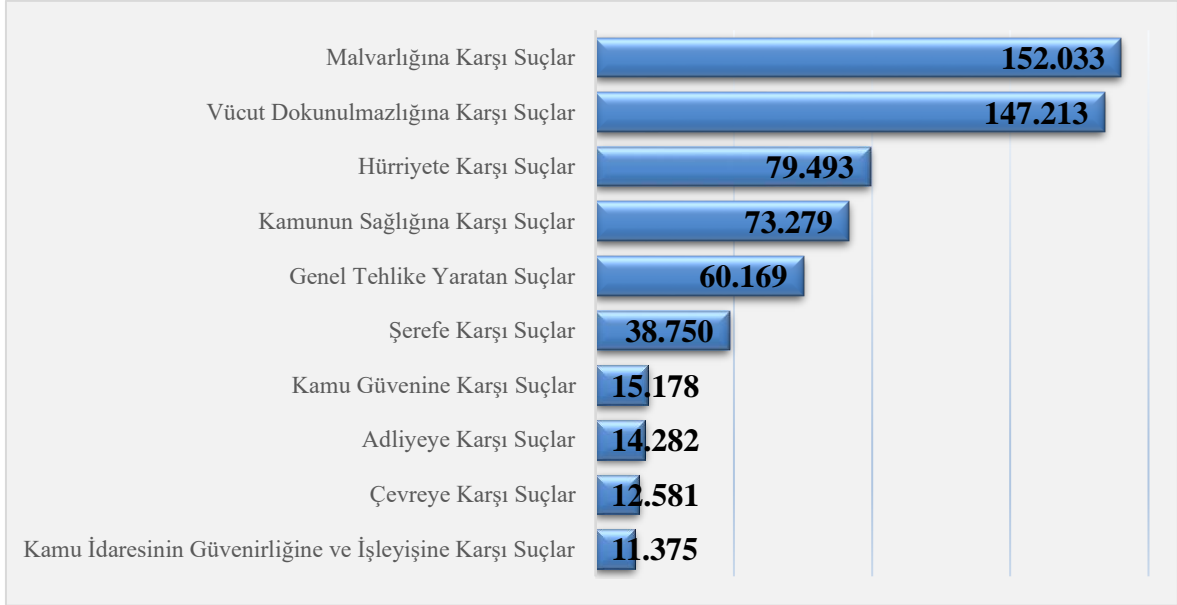


Şekil 3.1. TCK göre kısımlar ve bölümlere ait hiyerarşi



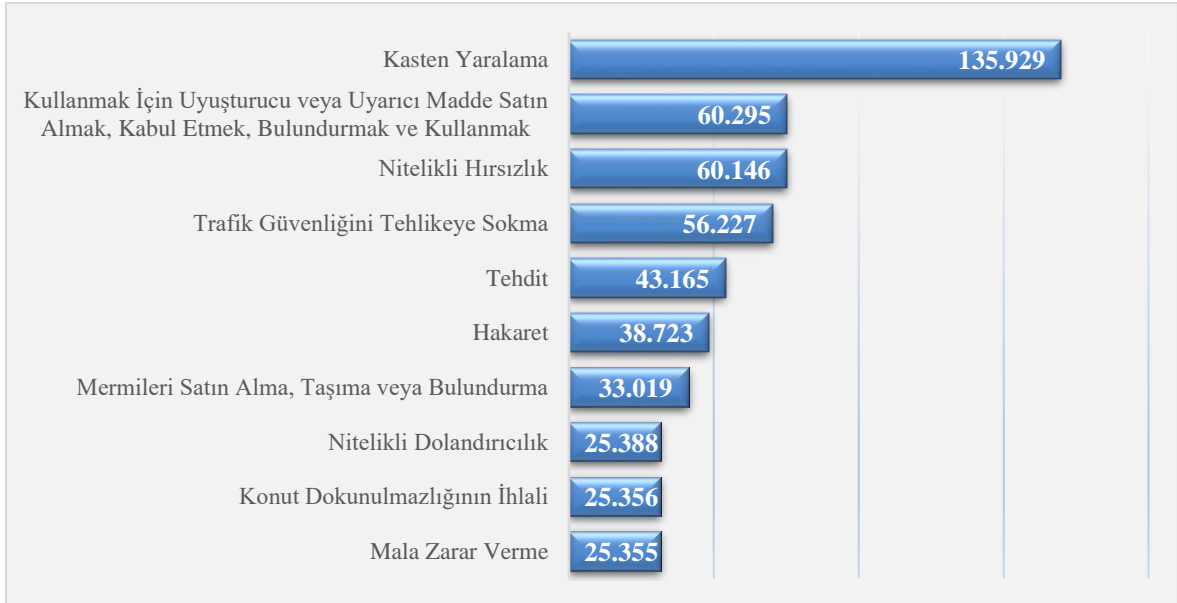
Şekil 3.2. TCK kapsamında kısımlara ait suç verilerinin oranları

Şekil 3.2’de gösterildiği üzere, Türk Ceza Kanunu; kişilere karşı suçlar, topluma karşı suçlar, millete ve devlete karşı suçlar ile uluslararası suçlar olmak üzere dört ana kısımdan oluşmaktadır. Bu kısımlara ait suç verileri incelendiğinde, en yüksek oran %68 ile kişilere karşı işlenen suçlara, en düşük oran ise %0,5 ile uluslararası suçlara aittir.



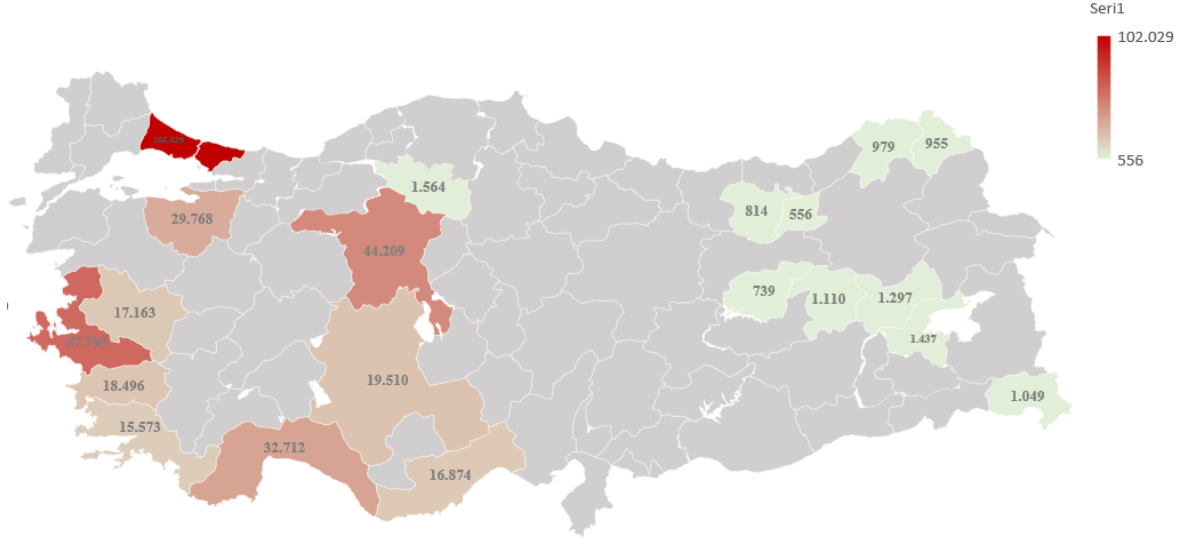
Şekil 3.3. TCK kapsamında bölümlere ait suç verilerinin frekansları

TCK'nın kısımları içerisinde yer alan bölümler incelendiğinde, ilk üç sırada “Kişilere Karşı Suçlar” başlığı altında yer alan malvarlığına karşı suçlar, vücut dokunulmazlığına karşı suçlar ve hürriyete karşı suçlar bulunmaktadır. Dördüncü sırada ise “Topluma Karşı Suçlar” kapsamında değerlendirilen kamunun sağlığına karşı suçlar yer almaktadır.



Şekil 3.4. Türkiye geneli suç türlerine ait ilk 10 suç

Türkiye geneli en çok işlenen suç türü ise kişilere karşı işlenen suçlar içerisinde yer alan kasten yaralama suçudur. 6136 sayılı kanun kapsamında yer alan mermileri satın alma, taşıma veya bulundurma suçu 7. Sırada yer almaktadır.



Harita 1.1. Suç kayıtlarına göre ilk 10 ve son 10 sıralarda yer alan iller

İl bazlı veriler analiz edildiğinde suç kaydı en fazla %14,7 oranında İstanbul'da iken %0,1 oranında Bayburt ilindedir. İlk 10 ile baktığımızda ise İstanbul ilini sırasıyla İzmir, Ankara, Antalya, Bursa, Konya, Aydın, Manisa, Mersin ve Muğla illeri takip etmektedir. Yukarıda yer alan Türkiye grafiğinde suç kaydı bakımından ilk ve son 10 suç kaydı bulunan iller yer almaktadır. Suç kaydının en yoğun olduğu iller kırmızı tonlarla, en az olduğu iller yeşil tonlarında gösterilmiştir.

Suç yoğunluğu yüksek olan iller genellikle kalabalık, karmaşık sosyal yapıya sahip, göç alan ve ekonomik çeşitliliği yüksek yerlerdir. Suç oranının düşük olduğu iller ise küçük, homojen, sosyal kontrolün güçlü olduğu ve göç almayan bölgeler olma eğilimindedir.



Harita 1.2. Suç kayıtlarına göre illerin yoğunlukları

Benzer biçimde Türkiye geneline baktığımızda suç kaydının en yoğun olduğu (kırmızı tonları) illerden en az olduğu (sarı tonları) illere göre renklendirilmiştir.

Suç ve cinsiyet dağılımı incelendiğinde, suç işleyenlerin %93,7'sinin erkek, %6,3'ünün ise kadın olduğu görülmektedir. Cinsiyet, bireylerin suça yönelme biçimlerini etkileyen önemli sosyo-demografik faktörlerden biridir.

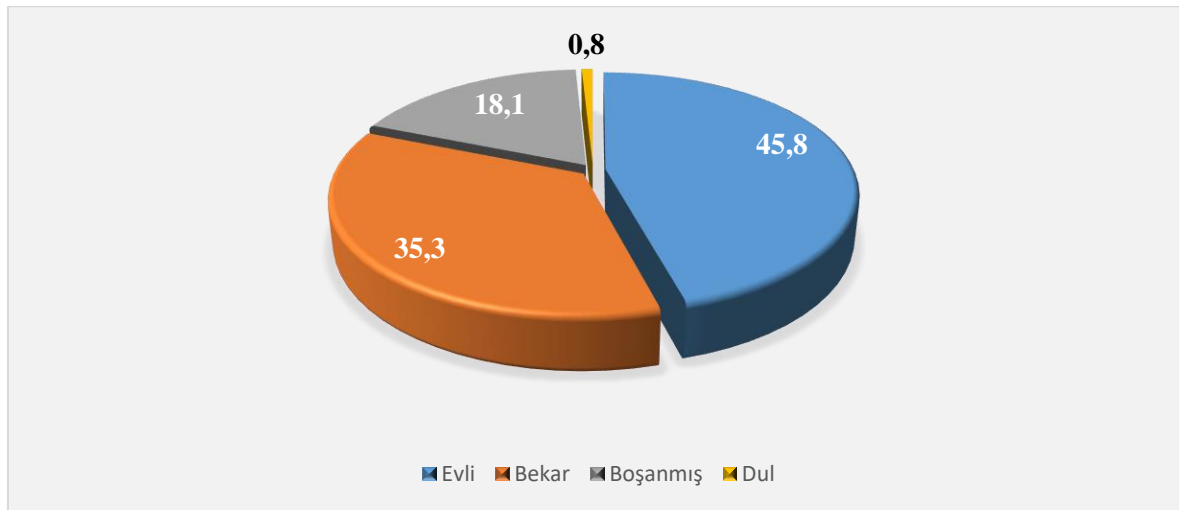
Erkekler için ilk 10 suç türü kasten yaralama, kullanmak için uyuşturucu veya uyarıcı madde satın almak, kabul etmek, bulundurmak ve kullanmak, nitelikli hırsızlık, trafik güvenliğini tehlikeye sokma, tehdit, hakaret, mermileri satın alma, taşıma veya bulundurma, nitelikli dolandırıcılık, konut dokunulmazlığının ihlali ve mala zarar verme suçlarıdır.

Kadınlar için ise ilk 10 suç türü kasten yaralama, nitelikli hırsızlık, hakaret, imar kirliliğine neden olma, tehdit, kullanmak için uyuşturucu veya uyarıcı madde satın almak, kabul etmek, bulundurmak ve kullanmak, mala zarar verme, konut dokunulmazlığının ihlali, trafik güvenliğini tehlikeye sokma ve nitelikli dolandırıcılık suçlarıdır.

Erkekler ve kadınlar arasında gözlemlenen suç türü farklılıkları, yalnızca bireysel eğilimlerle değil; toplumsal roller, sosyo-ekonomik koşullar ve kültürel normlarla da şekillenmektedir. Toplumda erkekler genellikle daha saldırgan, baskın ve dışa dönük rollerle özdeşleştirilirken; kadınlar daha itaatkâr, pasif ve ev içi rollerle ilişkilendirilir. Bu farklı

sosyalleşme süreci, suç işleme biçimlerine de yansımaktadır. Erkekler genellikle fiziksel şiddet, tehdit, silah bulundurma gibi daha dışa dönük ve kamusal alanda işlenen suçlara yönelirken; kadınlar daha çok dolaylı ifade biçimleri içeren suçlar (örneğin hakaret, imar kirliliği gibi) ile ilişkilendirilmektedir. Erkeklerin iş gücü piyasasında, sokakta ve kamusal yaşamda daha fazla yer alması, suç işleme fırsatlarını da artırmaktadır. Uyuşturucu madde temini, trafik suçları, silah bulundurma ve organize suçlar gibi eylemlere erkeklerin daha sık karışması bu durumla ilişkili olabilir. Kadınlar ise daha çok ev içi çevrelerde suç işlemektedir; bu da konut dokunulmazlığının ihlali ya da mala zarar verme gibi suçlarda görünürlük kazanmalarına neden olmaktadır. Kadınların iş gücüne katılım oranının düşüklüğü, ekonomik bağımsızlıklarının sınırlı olması, bazı durumlarda suç işlemeyi bir geçim aracı hâline getirebilir. Özellikle nitelikli hırsızlık ve dolandırıcılık gibi ekonomik temelli suçlarda kadınların oranı, bu durumla açıklanabilir. Kadınların bazı suçlara daha az şüpheli olarak görülmesi ya da suç ortaklığı durumunda göz ardı edilmesi, istatistiksel görünürlüklerini azaltabilir. Öte yandan erkeklerin suça daha yatkın görülmesi, onların daha kolay tespit edilip kayda geçirilmesine neden olabilir. Bu durum, verilerdeki oran farklılıklarını etkileyebilir.

Cinsiyete göre iller bazında dağılıma bakıldığında; erkekler açısından en yüksek suç oranlarının sırasıyla İstanbul, İzmir, Ankara, Antalya, Bursa, Konya, Aydın, Manisa, Mersin ve Gaziantep illerinde görüldüğü tespit edilmiştir. Kadınlar açısından ise en yüksek oranlara sahip ilk 10 il; İstanbul, İzmir, Ankara, Antalya, Bursa, Muğla, Aydın, Mersin, Denizli ve Konya'dır.



Şekil 3.5. Suç kaydı bulunan şahıslara ait medeni durum oranları

2023 yılına ait veriler incelendiğinde, suç kaydı bulunan bireylerin medeni durumlarına göre dağılımında evli bireylerin %45,8 ile en yüksek orana sahip olduğu görülmektedir. Evli bireyleri sırasıyla %35,3 oranıyla bekâr bireyler ve %18,1 oranıyla boşanmış bireyler takip etmektedir. En düşük oran ise %0,8 ile dul bireylerde gözlemlenmiştir.

Bu veriler, evli bireylerin mutlak sayı olarak suç kayıtlarında daha fazla temsil edildiğini göstermektedir. Ancak, bu dağılımın yorumlanmasında her bir medeni durum grubunun toplum içindeki genel nüfus oranlarıyla karşılaştırılması da önem arz etmektedir. Aksi takdirde, suç işleme eğiliminin medeni durumla doğrudan ilişkili olduğu yönünde yanıltıcı sonuçlara varılabilir.

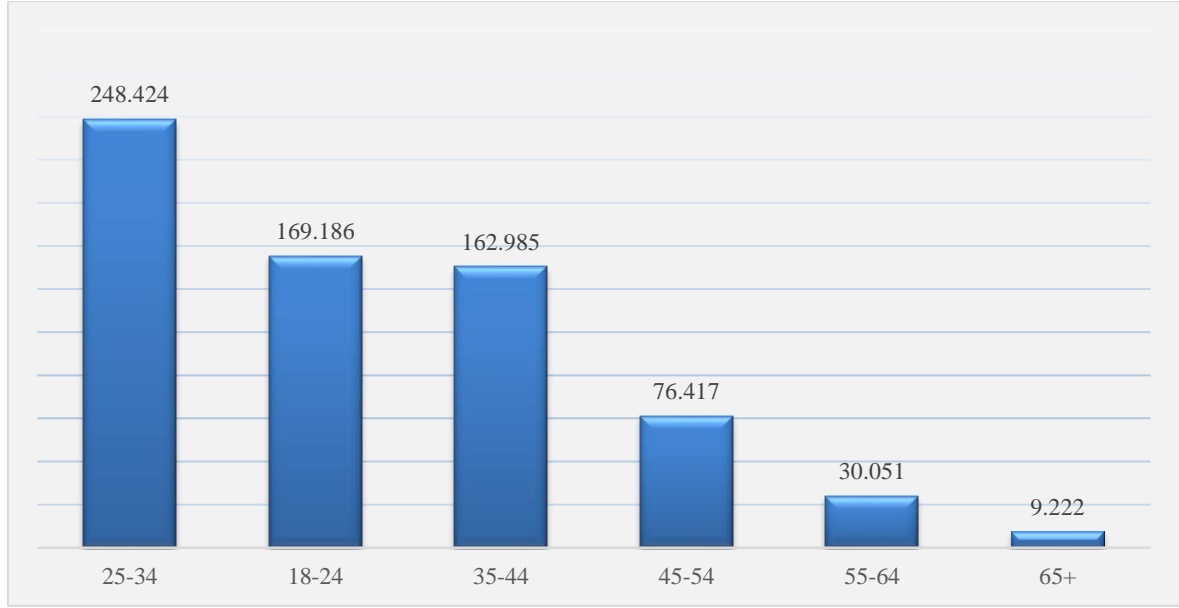
Bununla birlikte, suç kaydı bulunan bireylerin medeni durumuna göre dağılımında evlilerin en yüksek oranda yer almasının ya da dul bireylerin en düşük oranı oluşturmasının arkasında çeşitli sosyal, psikolojik ve demografik nedenler (toplumsal nüfus dağılımı, yaş faktörü ve suç eğilimi, aile ve ekonomik baskılar gibi) olabilir.

Evli bireylerde en çok rastlanan suç türleri arasında kasten yaralama, trafik güvenliğini tehlikeye sokma, tehdit, hakaret, uyuşturucu madde kullanımı ve bulundurulması, nitelikli hırsızlık, mermileri satın alma ve bulundurma, imar kirliliğine neden olma, kaçakçılık suçları ile mala zarar verme yer almaktadır. Bu gruba özgü olarak kaçakçılık suçları diğer medeni durumlara göre daha belirgindir.

Bekâr bireylerde ise nitelikli hırsızlık en sık işlenen suç türü olarak öne çıkmaktadır. Bunun yanı sıra kasten yaralama, uyuşturucu madde kullanımı ve bulundurulması, trafik güvenliğini tehlikeye sokma, konut dokunulmazlığının ihlali, mermileri satın alma ve bulundurma, mala zarar verme, nitelikli dolandırıcılık, tehdit ve hakaret suçları yaygın olarak görülmektedir.

Boşanmış bireylerde en çok rastlanan suçlar arasında kasten yaralama, tehdit, trafik güvenliğini tehlikeye sokma, uyuşturucu madde kullanımı ve bulundurulması, hakaret, nitelikli hırsızlık, nitelikli dolandırıcılık, mermileri satın alma ve bulundurma, mala zarar verme ile konut dokunulmazlığının ihlali yer almaktadır. Bu grupta nitelikli dolandırıcılık ve konut dokunulmazlığının ihlali suçları bekârlara benzerlik göstermektedir.

Dul bireylerde ise kasten yaralama, imar kirliliğine neden olma, hakaret, tehdit, trafik güvenliğini tehlikeye sokma, nitelikli hırsızlık, hakkı olmayan yere tecavüz, mala zarar verme, mermileri satın alma ve bulundurma ile uyuşturucu madde kullanımı ve bulundurma suçları görülmektedir. Dul bireylere özgü olarak hakkı olmayan yere tecavüz suçu sadece bu grupta yer almakta, ayrıca imar kirliliğine neden olma suçu da diğer gruplara kıyasla daha ön plandadır.



Şekil 3.6. Suç kaydı bulunan bireylere ait yaş gruplarına ilişkin frekanslar

Suç kaydı bulunan bireyler yaş gruplarına göre incelendiğinde, en yüksek oran %35,7 ile 25-34 yaş grubuna aittir. Bunu %24,3 ile 18-24 yaş grubu ve %23,4 ile 35-44 yaş grubu takip etmektedir. Diğer yaş gruplarına bakıldığında, 45-54 yaş grubunun oranı %11, 55-64 yaş grubunun oranı %4,3, 65 yaş ve üzeri bireylerin oranı ise %1,3'tür.

Tüm yaş gruplarında öne çıkan en yaygın suç türü kasten yaralamadır. Bu suç hem gençlerde hem de ileri yaşta kişilerde yüksek oranlarla ilk sıralarda yer almakta, şiddet içerikli suçların her yaş grubunda önemli bir problem olduğunu göstermektedir. Yaş ilerledikçe suç türlerinde değişiklikler gözlenmektedir. Genç yaşlarda fiziksel şiddet, hırsızlık ve madde kullanımı ön plandayken; orta yaşlarda sosyal ilişkilerle bağlantılı tehdit ve hakaret gibi suçlar öne çıkmakta, ileri yaşlarda ise çevre, mülkiyet ve imarla ilgili suçlar daha sık görülmektedir. Bu durum, bireylerin yaşadıkları hayat evrelerinin, suç davranışlarının doğasını etkilediğini göstermektedir.

18-24 yaş grubunda; kasten yaralama (%16,81), nitelikli hırsızlık (%14,81), uyuşturucu kullanımı (%11,20) ilk sıralarda yer almaktadır. Genç yaş grubunda özellikle mala karşı işlenen suçlar ile uyuşturucu kullanımı dikkat çekmektedir. Bu, riskli davranışların ve bağımlılıkla ilgili suçların gençlerde daha yaygın olduğunu göstermektedir.

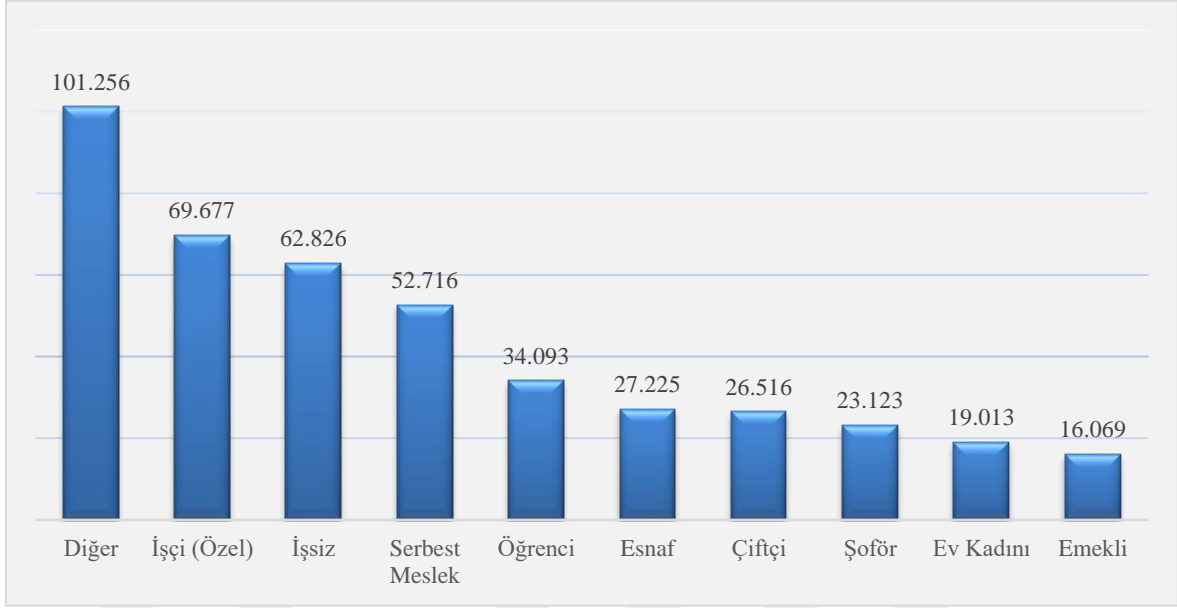
25-34 yaş grubunda; kasten yaralama (%19,04), uyuşturucu kullanımı (%11,68) ve nitelikli hırsızlık (%9,22) ön plana çıkmaktadır. Genç yetişkinlik döneminde şiddet ve madde kullanımı yine üst sıralarda yer almakta, aynı zamanda trafik güvenliğini tehlikeye sokma gibi davranışsal suçların da oranı artmaktadır.

35-44 yaş grubunda; kasten yaralama (%21,98), trafik güvenliğini tehlikeye sokma (%10,12) ve tehdit (%7,74) suçları öne çıkmaktadır. Orta yaşlara gelindikçe suç profilinde değişim yaşanmakta, şiddet suçlarının yanı sıra sosyal ilişkilerden kaynaklanan tehdit ve hakaret gibi suçlar da ön plana çıkmaktadır.

45-54 yaş grubunda; kasten yaralama (%21,67), trafik güvenliğini tehlikeye sokma (%11,81) ve hakaret (%8,71) suçları yüksek oranlara sahiptir. Orta yaşın ilerleyen dönemlerinde yine şiddet suçları devam ederken, çevreye ve kamuya zarar verme potansiyeli taşıyan suçların (örneğin imar kirliliği) da bu yaş grubunda öne çıktığı görülmektedir.

55-64 yaş grubunda da; kasten yaralama (%19,98), trafik güvenliğini tehlikeye sokma (%10,95) ve hakaret (%9,99) bu yaş aralığında en yaygın suçlardır. İleri yaşlara yaklaşan bireylerde özellikle hakaret ve imar düzenine aykırı faaliyetler gibi daha çok hukuki anlaşmazlık kaynaklı suçlar da öne çıkmaktadır.

65 yaş ve üzerinde yaşlı bireylerde en sık karşılaşılan suçlar; kasten yaralama (%19,32); imar kirliliğine neden olma (%14,92) ve hakkı olmayan yere tecavüzdür (%10,44). Bu yaş grubunda doğrudan fiziksel eylemlerden çok, mülkiyet ve çevreyle ilgili suçlar ön plana çıkmaktadır. Bu durum, yaşlı bireylerin daha çok tapu, arazi ve imar sorunları gibi konularda suçlarla ilişkilendirildiğini göstermektedir.



Şekil 3.7. Suç kaydı bulunan bireylere ait ilk 10 meslek gruplarına ilişkin frekanslar

Suç kaydı bulunan bireylerin meslek gruplarına bakıldığında, en yüksek oranın diğer (meslek gruplarına dâhil edilmeyen veya kategorize edilemeyen grupların) grubunda olduğu görülmektedir. İşsiz olanlar ve özel sektörde işçi olarak çalışanların suç işleme oranlarının yüksek olması, bu kişilerin içinde buldukları ekonomik ve sosyal zorlukların (sosyo-ekonomik kırılganlık) ile çalışma koşullarının (örneğin, düşük ücret, iş güvencesinin olmaması, stresli ortamlar gibi) suç davranışlarını artırıcı bir etkisi olduğunu gösterebilir.

Bu durum, işsizliğin suç işleme riski üzerinde önemli bir etkisi olduğunu göstermektedir. İşsizlik, bireylerin ekonomik ve sosyal sorunlarla karşılaşması nedeniyle suç davranışlarını artıran bir faktör olarak değerlendirilebilir.

Grafikte yer alan öğrenci ve ev hanımı kategorileri ise suçun sadece çalışma hayatıyla sınırlı olmadığını göstermektedir. Öğrenciler, genç nüfusun suç oranlarındaki etkisini ortaya koyarken, ev hanımlarında görülen suç kayıtları farklı sosyal dinamiklerin de suç üzerinde etkili olduğunu düşündürmektedir.

Meslek gruplarına göre işlenen suç türleri incelendiğinde, özel sektörde işçi olarak çalışanlarda kasten yaralama, uyuşturucu madde kullanımı ve trafik güvenliğini tehlikeye sokma suçlarının diğer gruplara göre daha yaygın olduğu görülmektedir. Bu durum, söz konusu bireylerin çalışma koşulları ve sosyal çevrelerinin davranışları üzerinde belirli ölçüde etkili olabileceğini düşündürmektedir. Benzer şekilde, işsizler, serbest meslek

sahipleri ve öğrenciler arasında kasten yaralama, nitelikli hırsızlık ve uyuşturucu madde kullanımını suçlarının öne çıkması, bu grupların sosyal çevreleri ve destek mekanizmalarının bazı zorluklarla karşılaşabileceğine işaret etmektedir.

Esnaflar ve çiftçiler arasında kasten yaralama, trafik güvenliğini tehlikeye sokma ve tehdit suçlarının nispeten daha sık görülmesi, bu meslek gruplarının ekonomik faaliyetlerinin getirdiği stres ve sosyal etkileşimlerin bazı zorluklarla karşılaşabileceğine işaret edebilir. Özellikle esnaflar arasında mermi satın alma, taşıma veya bulundurma suçlarının görece yüksek olması, bu meslek grubunda silah kullanımına dair belirli risklerin varlığını düşündürmektedir.

Şoförler arasında da kasten yaralama, trafik güvenliğini tehlikeye sokma ve uyuşturucu madde kullanımını suçlarının yaygın olması, mesleğin doğası gereği trafik kurallarıyla ilgili suçların daha sık işlendiğini göstermektedir. Ayrıca, iş ortamı ve yolculuk sırasında maruz kalınan stres faktörlerinin şiddet eğilimlerini artırabileceği değerlendirilebilir.

Ev hanımlarında yine kasten yaralama en fazla kasten yaralama suç türü yer alırken; nitelikli hırsızlık ve hakaret suçları diğer suç türleridir. Ev hanımlarında en yaygın suç türünün kasten yaralama olması, aile içi şiddet veya benzeri sosyal sorunların olabileceğini düşündürmektedir.

Emeklilerde ise kasten yaralama, trafik güvenliğini tehlikeye sokma ve hakaret suç türleri ilk üç sırada yer almaktadır.

Genel olarak, meslek gruplarına göre suç profilleri, sosyo-ekonomik durum, yaş, cinsiyet ve sosyal çevrenin suç davranışları üzerindeki etkisini kapsamlı bir şekilde ortaya koymaktadır. Bu bulgular, suçla mücadelede çok boyutlu ve hedef odaklı politikaların geliştirilmesinin önemini vurgulamaktadır. Özellikle işsizliğin azaltılması, sosyal destek mekanizmalarının güçlendirilmesi ve yüksek risk gruplarına yönelik özel önleyici programların hayata geçirilmesi suç oranlarının düşürülmesinde etkili olacaktır.

3.3. CC Algoritması Uygulaması

Satırlarda suç türlerinin sütunlarda illerin yer aldığı veri setine CC algoritması uygulanmıştır. Analiz yapılırken algoritmaya ait parametre değerlerinden kabul edilen maksimum H skorunu temsil eden delta değeri (δ) 0,05 alınırken, ölçek faktörünü temsil eden α değeri 1 olarak alınmıştır. Aynı zamanda bir diğer parametre olan maksimum ikili küme sayısının en fazla 50 adet olarak belirlenmiştir.

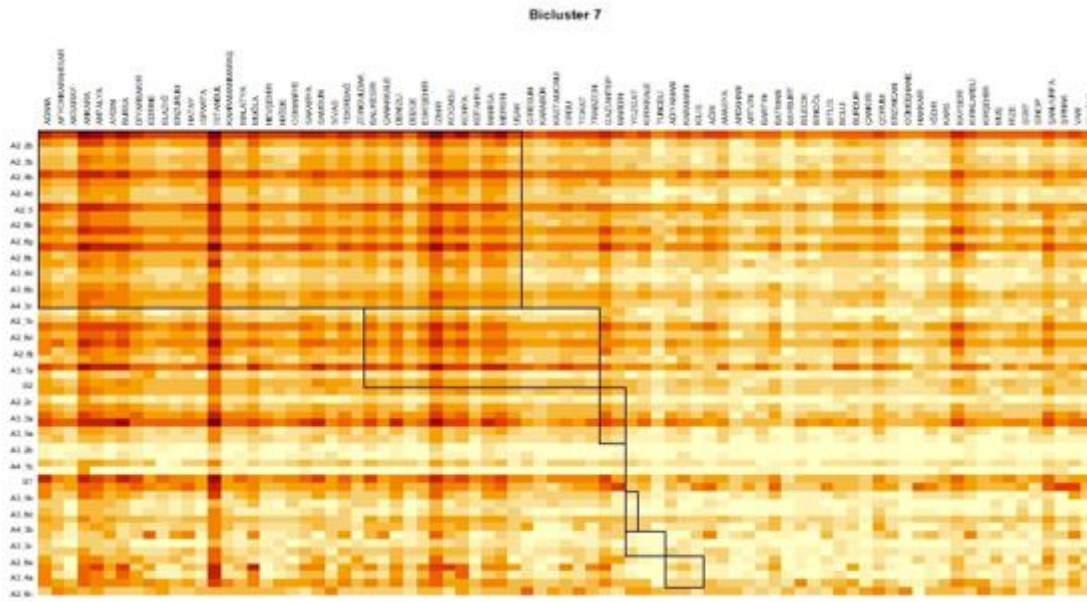
```
biclust_result_CC <- biclust(sucturu_log, method = BCCCO, delta = 0.05, alpha = 1, number = 50)
biclust_result_CC
```

Şekil 3.8. CC algoritmasına ait R kodu

Alınan bu parametrelere göre 7 adet ikili küme bulunmuştur. Aşağıda yer alan Çizelge 3.4'te ikili kümeye ait boyutlar yer almaktadır. İkili kümeleri büyüklükleri bakımından incelediğimizde İkili Küme 1'in 58 suç türünden 22'sini (%38,6) kapsarken, en küçük ikili küme de 3 suç türünü kapsayan İkili Küme 6'dır (%5,7).

Çizelge 3.4. CC algoritması ikili kümelerin boyutları

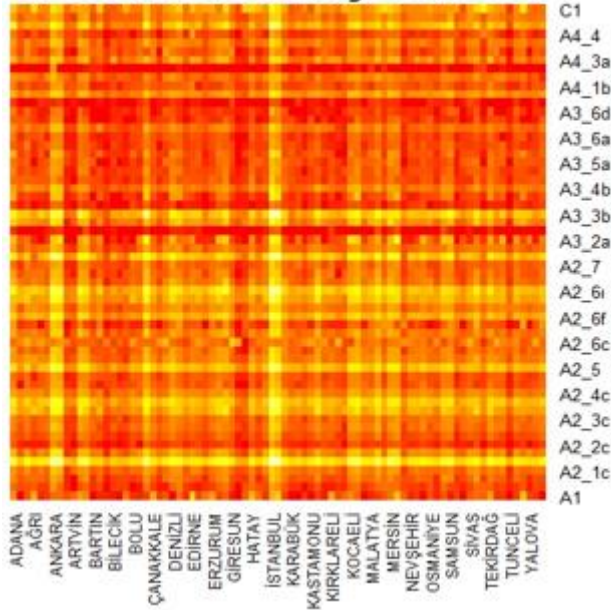
	BC1	BC2	BC3	BC4	BC5	BC6	BC7
Satır	22	10	7	6	5	3	4
Sütun	37	18	16	9	8	10	8



Şekil 3.9. CC algoritmasına ait ısı haritası

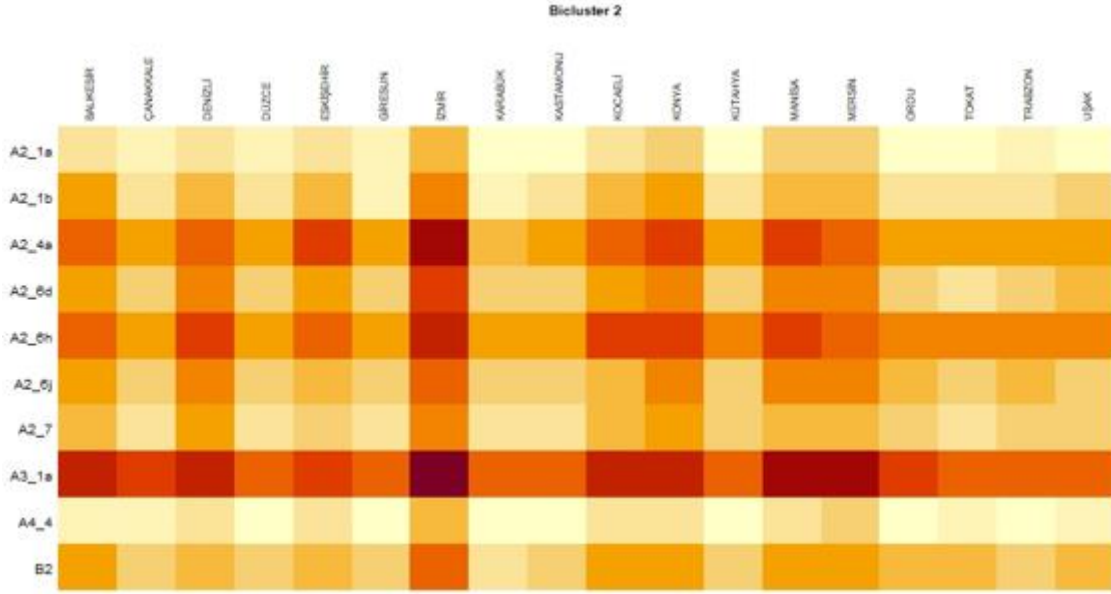
Yukarıda yer alan algoritmaya ait ısı grafiği yardımıyla bulunan 7 adet ikili kümeler gösterilmiştir. Bazı ikili kümeler için örtüşme durumu söz konusudur. Bir küme içerisinde yer alan bazı iller benzer istatistiklere sahip olması nedeniyle başka küme içerisinde de yer almıştır.

Normalize Edilmiş Veri Seti



Şekil 3.10. Veri matrisine ait ısı haritasının genel görünümü

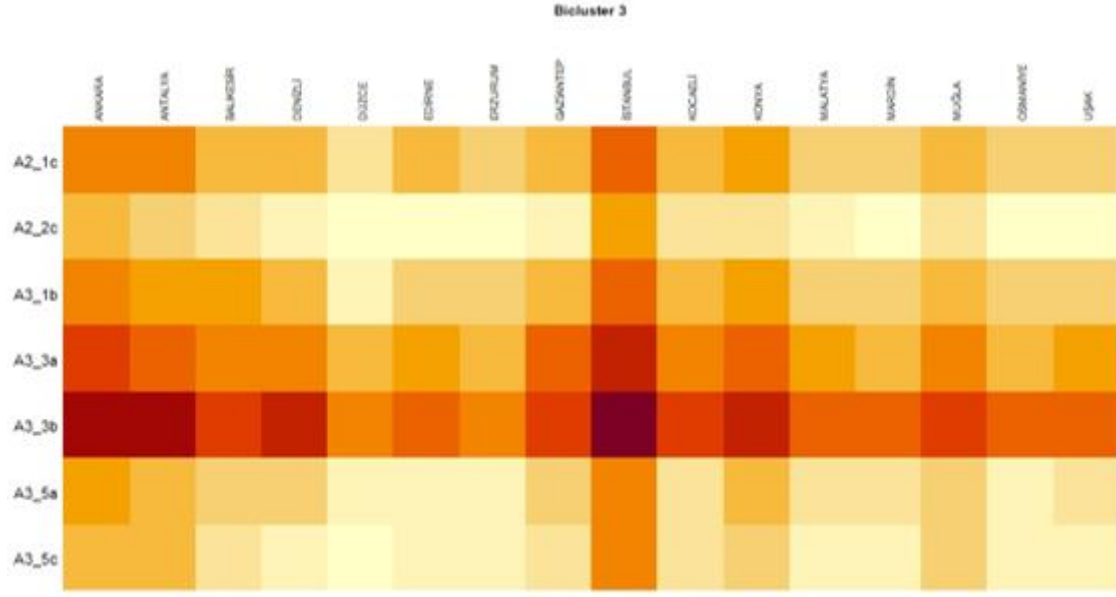
Şekil 3.4 ve Harita 1.1'de yer alan tanımlayıcı istatistiklerde de gösterilen en yoğun suç türlerinden ve en yoğun illeri aynı kümede bir araya getirmiştir.



Şekil 3.12. CC algoritması ikili küme 2'ye ait ısı haritası

İkili küme 2'ye ait olan ısı grafiğinde ise topluma karşı işlenen suçlar yer alan genel tehlike yaratan suçlar içerisinde trafik güvenliğini tehlikeye sokma suçu (A3_1a) en yoğun olarak İzmir'de gözlemlenmektedir. Konut dokunulmazlığının ihlali (A2_4a) ve nitelikli dolandırıcılık (A2_6h) suç türleri de yoğunluk oranları fazla çıkan bir diğer suç türleridir.

Ayrıca grafikte yer alan suçlar kapsamında Çanakkale ve Bursa illeri, Balıkesir, Denizli ve Eskişehir illeri, Manisa ve Konya illeri ile Kocaeli ve Mersin illerinin suç profilleri benzerlik göstermektedir.

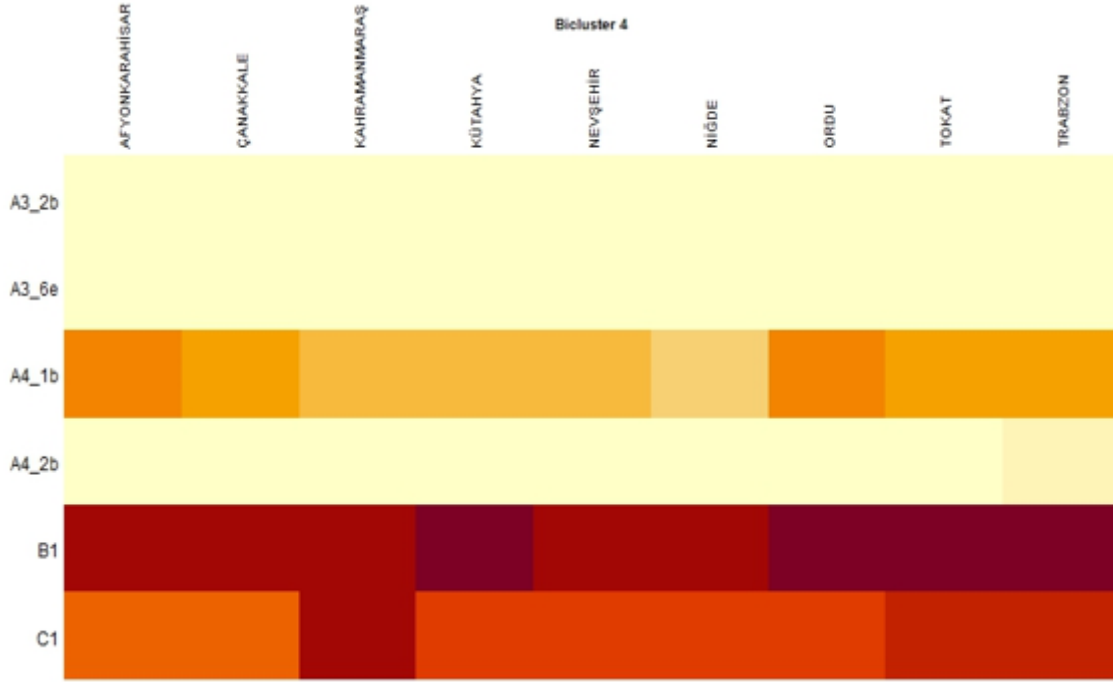


Şekil 3.13. CC algoritması ikili küme 3'e ait ısı haritası

İkili Küme 3'e ait ısı haritası incelendiğinde, suç türü bakımından kamunun sağlığına karşı işlenen suçlar kapsamında yer alan kullanmak için uyuşturucu veya uyarıcı madde satın alma, kabul etme ve bulundurma (A3_3b) suçunun bu kümede en yoğun şekilde öne çıktığı görülmektedir. Bu suç türü, özellikle metropolitan bir şehir olan İstanbul'da yüksek oranda kümelenmiş olup, bu durum büyükşehirlerde uyuşturucu kullanımına bağlı suçların yaygınlığına dair genel eğilimle örtüşmektedir.

Bununla birlikte, Ankara, Antalya, Denizli ve Konya gibi gelişmiş ve görece yüksek nüfusa sahip illerde de bu suçun yoğunlaştığı gözlemlenmektedir. Bu durum, uyuşturucu kullanımı ve bulundurma suçunun yalnızca büyükşehirlerle sınırlı olmadığını, bölgesel düzeyde de önemli bir sorun alanı oluşturduğunu da göstermektedir.

Ayrıca, uyuşturucu veya uyarıcı madde imal ve ticareti (A3_3a) suçunun da bu kümede dikkat çekici bir şekilde yer aldığı görülmektedir. Bu suç türü de benzer şekilde İstanbul'da yoğun olarak kümelenmiş, ancak Ankara, Antalya, Gaziantep ve Konya illerinde de belirgin düzeyde öne çıkmıştır. Bu bulgu, yalnızca kullanım değil, aynı zamanda uyuşturucu madde ticareti açısından da belirli illerde risk düzeyinin yüksek olduğu dikkat çekmektedir.

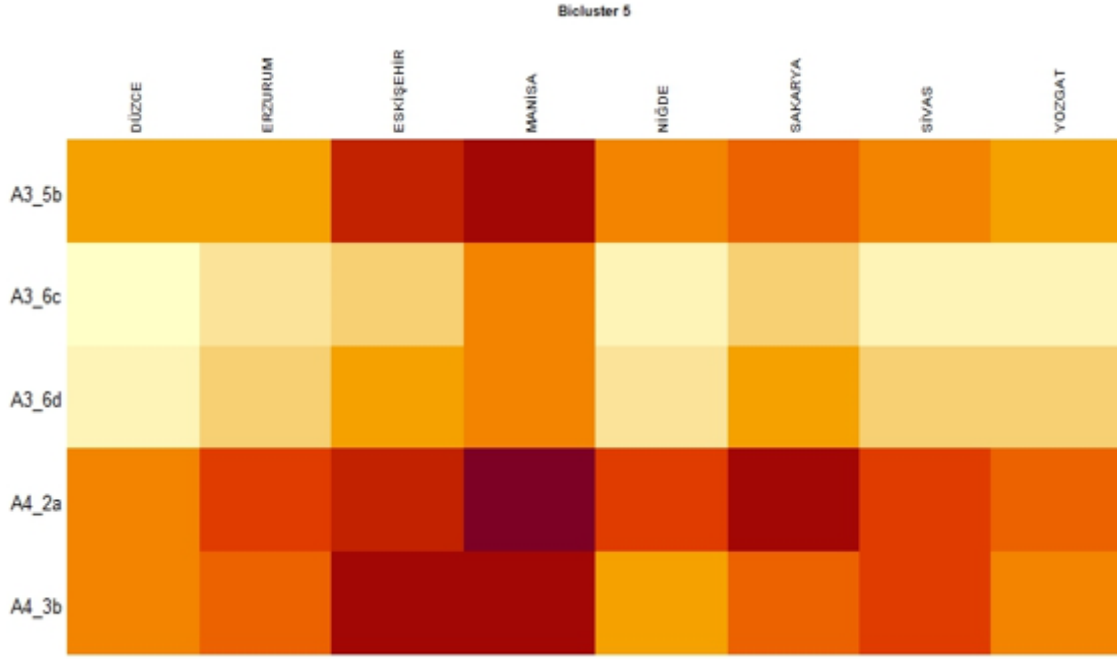


Şekil 3.14. CC algoritması ikili küme 4'e ait ısı haritası

İkili küme 4'e ait ısı grafiğinde 6136 sayılı kanun kapsamında yer alan mermileri satın alma, taşıma veya bulundurma suçu (B1) en yoğun suç türü olarak çıkmıştır. Özellikle Kütahya, Ordu, Tokat ve Trabzon'da daha baskın olduğu gözlemlenmektedir.

Yoğunluğu fazla olan bir diğer suç türü ise 5607 sayılı kanun olan kaçakçılık suçları (C1) oluşturmaktadır. Kahramanmaraş'ta en yoğun iken Tokat ve Trabzon illerinde de baskın olarak çıkmıştır.

İkili Küme 4'teki sonuçlar, ateşli silahlar ve kaçakçılıkla ilgili suçların coğrafi olarak belli illerde kümelenildiğini göstermektedir. Bu kümelenme; coğrafi konum, ulaşım ağlarına yakınlık, sosyoekonomik yapı ve güvenlik denetimlerinin yoğunluğu gibi faktörlerden etkilenmiş olabilir. Özellikle silah bulundurma ve kaçakçılık gibi suçlar, bireysel güvenlik kaygıları ya da ekonomik fırsatlar ile ilişkili olarak bazı bölgelerde daha yoğun şekilde ortaya çıkmaktadır.



Şekil 3.15. CC algoritması ikili küme 5'e ait ısı grafiği

İkili küme 5'e ait ısı grafiğinde ise millete ve devlete karşı suçlar kapsamında değerlendirilen silahlı örgüt (A4_2a) ve infaz ve tutukevine yasak eşya sokma (A4_3b) suçlarının özellikle Manisa'da daha yoğun olarak çıktığı gözlemlenmiştir. Ayrıca kumar oynamak için yer veya imkân sağlama (A3_5b) suç türü de Manisa'da daha yoğun ve Eskişehir'de de yoğun olarak çıkmıştır.

CC algoritmasının sonuçları genel olarak incelendiğinde, en fazla işlenen suç türleri ile bu suçların yoğun olarak kaydedildiği illerin birlikte kümelendiği görülmektedir. Özellikle ilk 10 suç türü ile en fazla suç kaydına sahip ilk 10 il, kümeleme kapsamında öne çıkan örüntüleri oluşturmuştur.

Bu durum hem algoritmanın etkinliğini göstermekte hem de tanımlayıcı istatistiklerle uyumlu bir şekilde örüntüler ortaya koyduğunu işaret etmektedir. Elde edilen kümeler, suçların belli coğrafi bölgelerde yoğunlaştığını ve bazı suç türlerinin belirli illerde daha sık işlendiğini açıkça ortaya koymuştur. Bu bulgular, kümeleme algoritmasının gerçek veriler üzerinde anlamlı sonuçlar ürettiğini ve istatistiksel analizle desteklenen bir yapı sunduğunu göstermektedir.

3.4. xMotif Algoritması Uygulaması

CC algoritması ile aynı değişkenler kullanılan ve suç türlerinin düşük, orta ve yüksek seviyelerine göre sınıflandırılarak kategorik hâle getirilen veri seti üzerinde xMotif algoritması uygulanmıştır.

xMotif algoritması satırları "örnekler", sütunları "özellikler" olarak kabul eder. Parametre içerisinde yer alan satır örnek sayısı (number of row samples-ns), her iterasyonda kaç satır (örnek/il) üzerinden alt kümeleme yapılacağını belirlerken; sütun örnek sayısı (number of column samples-nd) her iterasyonda kaç sütun (özellik/suç türü) üzerinden analiz yapılacağını belirler. Sütun özellik sayısı (minimum number of columns – sd) parametresi ise bir ikili kümenin geçerli sayılabilmesi için içermesi gereken en az sütun (özellik) sayısını ifade etmektedir. 58x81 boyutunda olan veri matrisinde parametre değerleri olarak ns ve nd değerleri suç türlerinin %34'ünü, illerin %25'ini kapsayacak şekilde 20 olarak alınmıştır.

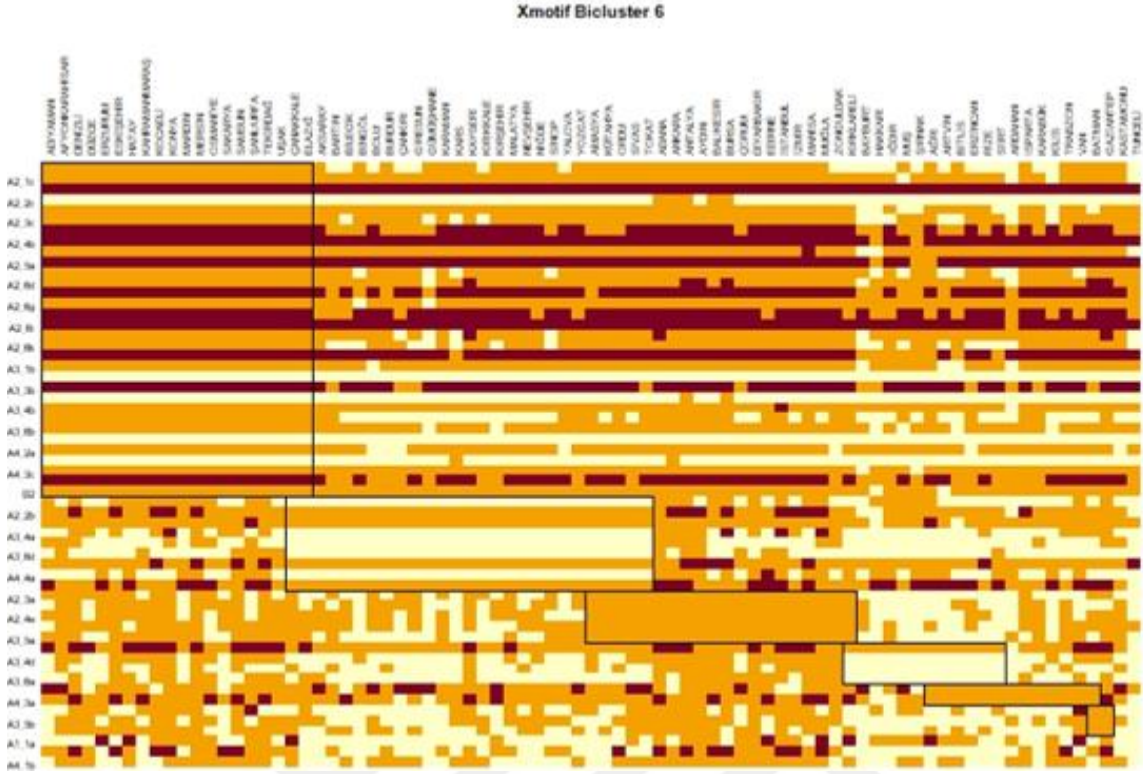
```
res <- biclust(discrete_data, method=BCXmotifs(), ns=20, nd=20, sd=5, alpha=0.05, number=5)
res
```

Şekil 3.16. xMotif algoritması için kullanılan R program kodları

Yukarıda yer alan parametreler sonucunda bulunan 6 adet ikili kümelerin boyutları Çizelge 3.5'te yer almaktadır. En büyük ikili küme 1 iken en küçük ikili küme 5'tir.

Çizelge 3.5. xMotif algoritması ikili kümelerin boyutları

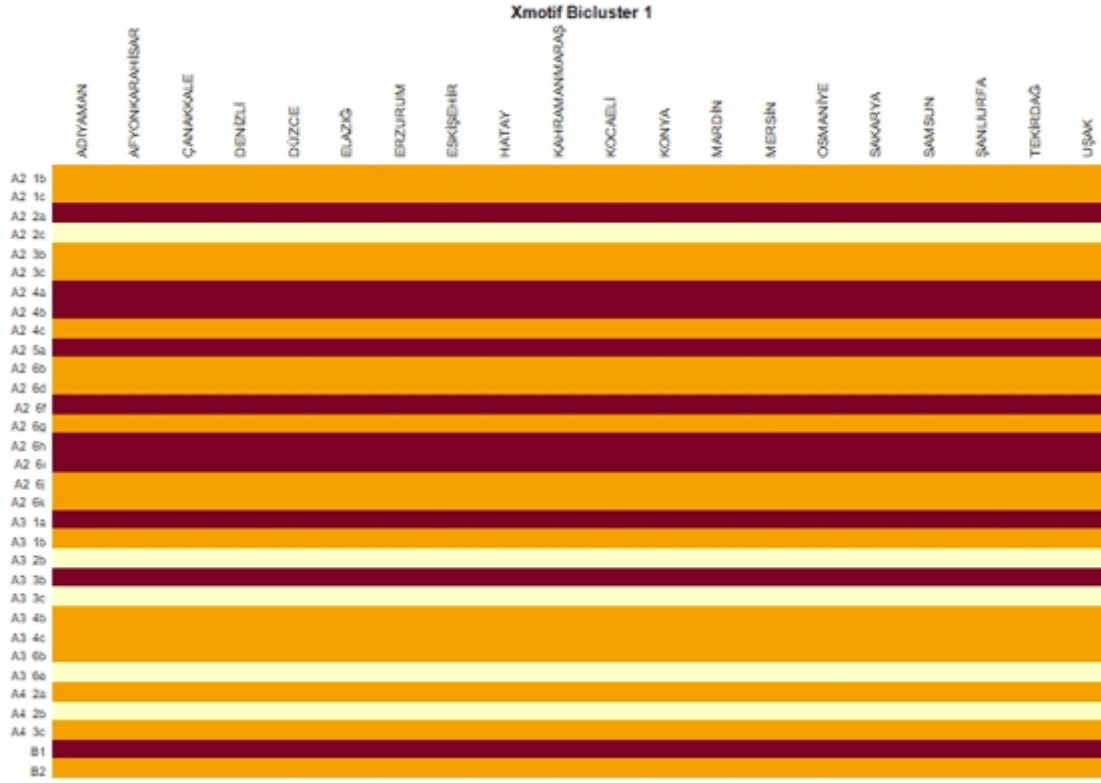
	BC1	BC2	BC3	BC4	BC5	BC6
Satır	32	9	5	4	2	3
Sütun	20	27	29	28	41	18



Şekil 3.17. xMotif algoritmasına ait ısı haritası

Kategorik yapıda olan veri setiyle oluşturulan ısı haritasında, verinin sınırlı değer aralığına sahip olması nedeniyle satırlar şeritler halinde gruplanmış olarak görünmektedir. Bu durum, özellikle kategorik verilerde benzer değerlerin tekrarlamasından kaynaklanmaktadır.

xMotif algoritması sonucunda elde edilen ikili kümeler arasında örtüşme (overlapping) gözlemlenmiştir. Bu, bazı illerin veya suç türlerinin birden fazla kümede yer aldığını, dolayısıyla bu kümelerin birbirini dışlamayan, ortak örüntüler içerdiğini göstermektedir. xMotif algoritmasının doğası gereği, örtüşen kümelerin ortaya çıkması beklenen ve kabul edilebilir bir durumdur.

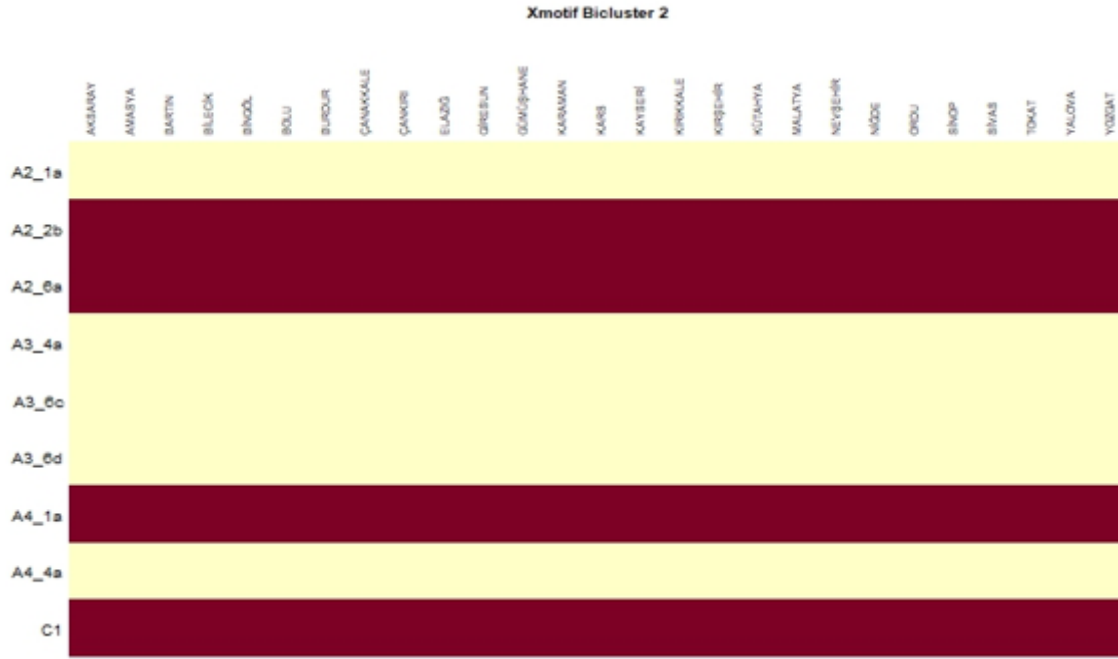


Şekil 3.18. xMotif algoritması ikili küme 1'e ait ısı haritası

Satırlarda yer alan suç türleri, sınıflandırmaya göre renk tonlarıyla görselleştirilmiştir. Yüksek düzeyde sınıflanan veriler koyu kırmızı, orta düzeydekiler turuncu ve düşük düzeydekiler sarı tonlarında gösterildiği gözlemlenmektedir.

Kasten yaralama, konut dokunulmazlığının ihlali, tehdit, hakaret, mala zarar verme, nitelikli dolandırıcılık, nitelikli hırsızlık, trafik güvenliğini tehlikeye sokma, kullanmak için uyuşturucu veya uyarıcı madde satın alma, kabul etme ya da bulundurma ile mermi satın alma ve bulundurma suçları, Şekil 3.4'te de görüldüğü üzere en yoğun kümelenen ilk 10 suç türünü oluşturmaktadır.

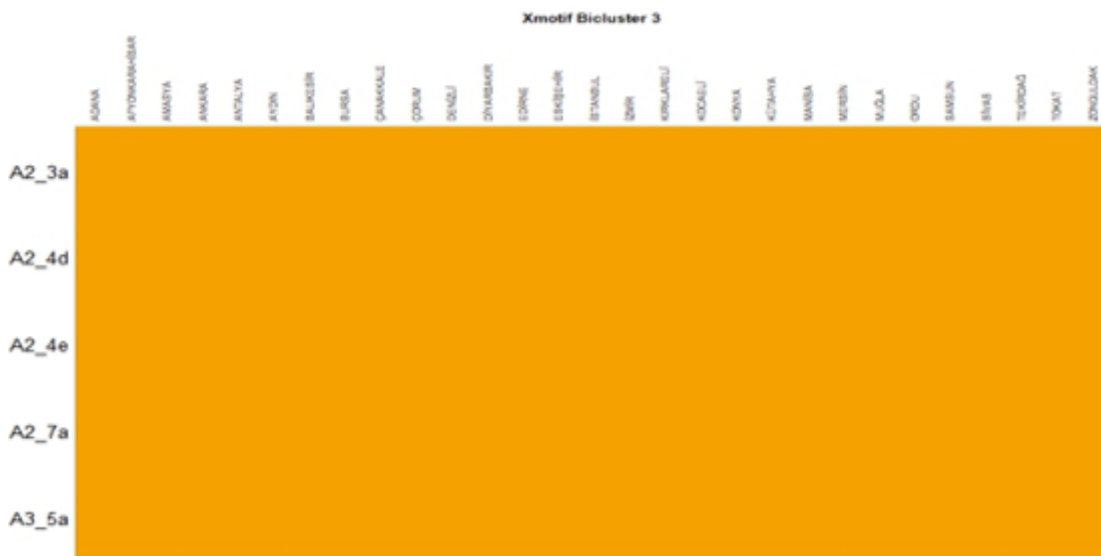
İller açısından değerlendirildiğinde ise, ortalaması 3 veya 3'e yakın olan iller benzer şekilde gruplanmıştır. Ancak, verilerin kesikli hale getirilmesi ve üç kategoriye ayrılması nedeniyle, frekansı daha yüksek olan bazı iller bu kümeler arasında yer almadığı düşünülmektedir.



Şekil 3.19. xMotif algoritması ikili küme 2'ye ait ısı haritası

Grafikte yer alan bu suç türleri, suç kaydı sayısı yüksek olan suçlar arasında yer almaktadır. Ayrıca, söz konusu kümede orta yoğunluktaki suç türlerine yer verilmemiş, kümelenme yalnızca yoğunluğu en yüksek ve en düşük olan suç türleri üzerinden gerçekleştirilmiştir.

Yukarıda yer alan ısı haritasında, taksirle yaralama, dolandırıcılık, görevini yaptırmamak için direnme ve kaçakçılık suçları, görece yüksek suç oranına sahip iller bakımından kümelenmiştir.



Şekil 3.20. xMotif algoritması ikili küme 3'e ait ısı haritası

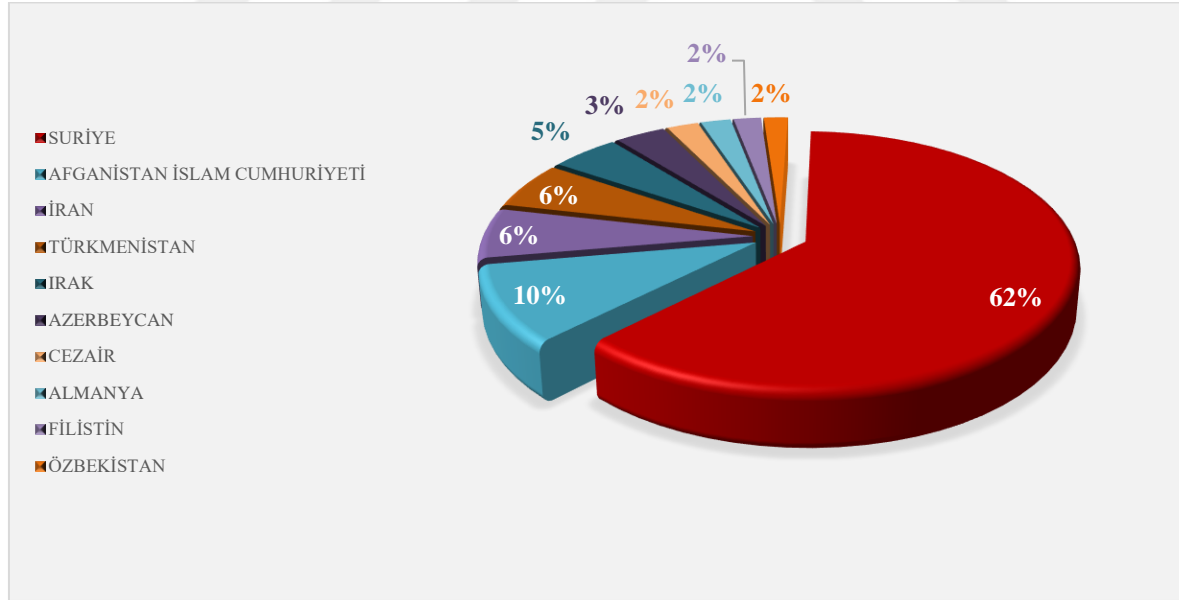
Kümeler belirli bir yoğunluk aralığına göre (örneğin sadece orta düzeydeki iller) oluşturulduğunda, aşırı yüksek değerlere sahip iller bu kümelere uyumsuz düşmekte ve dışarda kalmaktadır.

Kesikli veri, sayısal ayrıntıyı azalttığı için ince farklılıklar göz ardı edilmekte, bu da yüksek suç oranına sahip bazı illerin analizde temsil edilmemesine neden olmaktadır.

Bu nedenle, kesikli veri kullanımını analizde yorumlamayı kolaylaştırırsa da veri kaybı ve yüksek frekanslı illerin kümeler dışında kalması gibi sınırlamalar doğurabilmektedir.

3.5. Bimax Algoritması Uygulaması

Analizde, ikili değerlerden oluşan ve satırlarda uyrukların, sütunlarda ise suç türlerinin yer aldığı 100x58 boyutlarında bir matris kullanılmıştır. Veri seti, toplam 7.429 suç kaydını içermektedir. Bu yapı üzerinden uyruk ve suç türleri bakımından detaylı analiz gerçekleştirilmiştir.



Şekil 3.23. Suç kaydı yüksek olan ilk 10 uyruğun dağılım oranları

Yukarıda yer alan pasta grafiğinde de görüldüğü üzere Suriye uyruklular %62, suç işleyen yabancılar arasında açık ara en yüksek paya sahiptir. Bu durum, Suriyeli nüfusun hem sayısal olarak fazlalığını hem de bazı sosyal ve ekonomik sorunlara bağlı olarak suçta karışma oranlarının yüksek olabileceğini göstermektedir.

Afganistan uyruklular (%10) ile İran ve Türkmenistan uyruklular (%6'şar) da dikkate değer oranlara sahiptir. Irak (%5), Azerbaycan (%3) ve kalan ülkeler olan Cezayir, Almanya, Filistin ve Özbekistan (her biri %2) ile daha düşük oranlarda temsil edilmektedir.

Suriyeli ve Afgan göçmenlerin büyük çoğunluğu çatışma bölgelerinden geldiği için sosyoekonomik sorunlar, entegrasyon zorlukları, dışlanma gibi nedenler suça yönelimi artıran faktörler olabilir.

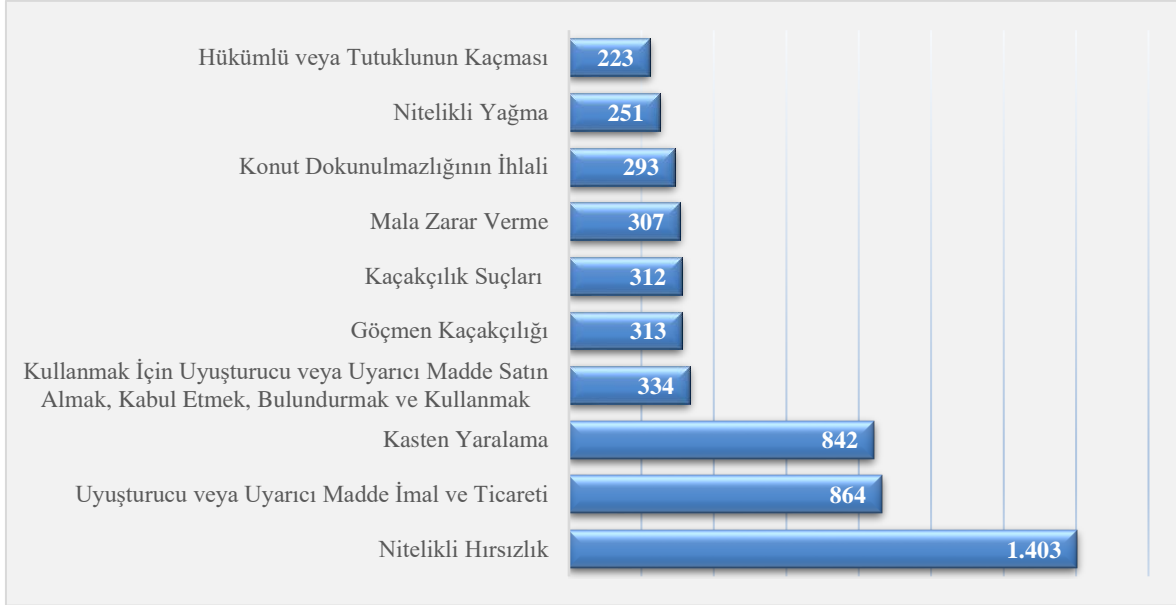
Almanya gibi görece refah düzeyi yüksek ülkelere gelen bireylerin oranı oldukça düşüktür; bu da göç nedeni ve göçmen profiliyle suç eğilimi arasında ilişki olabileceğini düşündürmektedir.

Suriye uyruklu kişilere ait suç kayıtlarına bakıldığında, en fazla kaydın İstanbul ilinde bulunduğu görülmektedir. İstanbul'u sırasıyla İzmir, Mersin, Gaziantep, Bursa, Hatay, Şanlıurfa, Kahramanmaraş, Adana ve Konya illeri takip etmektedir. Bu veriler, Suriyelilerin Türkiye'deki nüfus büyüklüğü ve yerleşim yoğunluğu, bu gruptaki suç oranlarının yüksekliğini açıklayan önemli bir etkidir.

Afganistan uyruklu şahıslar açısından değerlendirildiğinde, en fazla suç kaydının bulunduğu il yine İstanbul olmuştur. Bunu sırasıyla Nevşehir, Aksaray, Konya, Denizli, Eskişehir, Kayseri, Trabzon, Uşak ve Kırşehir izlemektedir. Bu durum, Afganistan uyrukluların yalnızca büyükşehirlerde değil, aynı zamanda İç Anadolu ve Karadeniz bölgelerindeki bazı orta ölçekli illerde de yoğunlaştığını ortaya koymaktadır.

İran uyruklu kişilere ait suç kayıtlarında da ilk sırada İstanbul yer almakta; ardından Antalya, Ağrı, Denizli, Erzurum, Van, Eskişehir, Kırşehir, Nevşehir ve Sivas illeri gelmektedir. Bu dağılım, İranlıların hem batıdaki turistik şehirlerde hem de doğu sınırına yakın illerde kayda değer bir varlık gösterdiğini ortaya koymaktadır.

Afgan ve İran uyruklularının bazı İç Anadolu ve Doğu Anadolu illerinde yoğunlaşması, göç rotaları ve barınma politikalarıyla ilişkilendirilebilir.



Şekil 3.24. Uyuşturuculara ait en çok işlenen ilk 10 suç türü

Suriyelilerde en sık rastlanan suç nitelikli hırsızlık ve kasten yaralamadır. Bu durum, büyük şehirlerdeki barınma ve geçim sorunlarıyla ilişkili olabilir. Ayrıca uyuşturucu ile ilgili suç oranı da dikkat çekicidir. Göçmen kaçakçılığı da önemli bir başlıktır, bu da sınır geçişleri ve insan ticaretiyle ilgili yapılarla ilişkilendirilebilir.

Afganlar arasında en yaygın suç türü uyuşturucu ticareti ve kasten yaralamadır. Aynı zamanda %4,8 ile çocuk istismarı gibi ciddi suçlar da sayısal olarak dikkat çekmektedir. Göçmen kaçakçılığı oranı da oldukça yüksektir, bu durum Afganistan'dan yasa dışı yollarla gelen bireylerin içinde bulunduğu kaçakçılık ağlarını işaret edebilir.

İran uyruklularda uyuşturucu suçları ilk sıradadır. İran'ın coğrafi konumu nedeniyle uyuşturucu güzergâhı üzerinde yer alması ve sınır geçişlerinin yoğunluğu bu durumu etkiliyor olabilir. Bunun yanı sıra resmî belgede sahtecilik ve trafik güvenliğini tehlikeye sokma gibi suçlar da göze çarpmaktadır.

Ekonomik nedenlerle Türkiye'ye gelen Türkmen vatandaşlarının en çok hırsızlık ve mala zarar verme gibi suçlara karışması, geçim sıkıntılarını yansıtıyor olabilir. Ayrıca uyuşturucu bağlantılı suçlar da dikkat çekmektedir.

Iraklılar arasında şiddet ve örgüt bağlantılı suçlar öne çıkıyor. Özellikle silahlı örgüt ve müstehcenlik gibi suçların yer alması güvenlik bağlamında dikkat çekicidir.

Azerbaycanlı bireylerin karıştığı suçlar daha çok şiddet ve hırsızlık eksenindedir. Fuhuş ve hakaret gibi bireyler arası suçların da dikkat çektiği görülüyor.

Cezayirliler arasında en yaygın suç nitelikli hırsızlıktır. Bu yüksek oran, bireysel ekonomik nedenlerin etkisini düşündürmektedir.

Türkiye'ye gelen Alman vatandaşlarının işlediği suçlar arasında şiddet ve çevre/imar suçları ön plandadır. Bu durum, tatil bölgelerindeki mülk sorunları ya da bireysel çatışmaları işaret ediyor olabilir.

Filistinliler arasında özellikle hırsızlık ve uyuşturucu suçları öne çıkıyor. Görece düşük sayılarla birlikte ciddi (hükümlü veya tutuklunun kaçması gibi) suç türleri görülmektedir.

Özbekistan uyruklular arasında şiddet ve tehdit suçları öne çıkıyor. Fuhuş ve dolandırıcılık gibi suç türleri, sosyal yapılarla ilgili farklı risk alanlarını işaret etmektedir.

İlk 10 uyruk arasında en yaygın suç türleri; nitelikli hırsızlık (hemen her uyrukta ilk 3'te), kasten yaralama ve şiddet suçları, uyuşturucu ticareti ve kullanımı, göçmen kaçakçılığı ile konut dokunulmazlığının ihlali ve mala zarar vermedir.

Bu durum, göçmenlerin barınma, işsizlik, sosyal dışlanma gibi sorunlar nedeniyle suça yönelme eğiliminde olduklarını göstermektedir. Özellikle sınır bölgelerinden giriş yapan ve büyükşehirlerde yaşayan gruplarda suç oranları artış göstermektedir.

Ayrıca bazı uyruklarda örgüt bağlantılı suçlar, cinsel istismar, fuhuş ve belge sahteciliği gibi spesifik suç türleri de dikkat çekmektedir. Bu suçlar, Türkiye'nin göç politikalarının, entegrasyon süreçlerinin ve güvenlik önlemlerinin ne kadar hayati olduğunu ortaya koymaktadır.

Uyruklar cinsiyet açısından değerlendirildiğinde, suç kayıtlarının büyük çoğunluğunun erkeklere ait olduğu görülmektedir. Ancak, Kırgızistan ve Ukrayna uyruklular arasında kadınlara ait suç kayıtlarının diğer uyruklara kıyasla daha yüksek olması dikkat çekmektedir.

Medeni durum bakımından değerlendirildiğinde, suç kaydı bulunan kişilerin %58,4'ünün bekâr, %39'unun evli, %2'sinin boşanmış ve %0,5'inin dul olduğu görülmektedir.

Türkiye’den farklı olarak, suç kaydı bulunan bekâr bireylerin oranının yüksek olmasının, göçmen profiline bağlı olarak açıklanabileceği değerlendirilmektedir. Nitekim göç eden bireyler arasında bekârların, evli bireylere kıyasla daha kolay ve esnek şekilde göç edebildiği, dolayısıyla bu grubun suç kaydı istatistiklerinde daha fazla temsil edildiği düşünülmektedir.

Uyrukların yaş gruplarına ait istatistiklerde 18-24 yaş grubu daha yüksek frekansa sahipken, 22 yaşındaki uyruklular ilk sırada yer almaktadır.

Uyruklara ait veriler eğitim durumu bakımından değerlendirildiğinde, suç kaydı bulunan kişiler arasında en yüksek oran %25,8 ile ilköğretim mezunlarına aittir. Bunu sırasıyla %16,8 ile okuma-yazma bilmeyenler, %16,3 ile ortaokul veya dengi meslek okulu mezunları, %15 ile lise veya dengi meslek okulu mezunları, %9,9 ile ilköğretim mezunları, %9,5 ile okur-yazar olup herhangi bir okul bitirmemiş bireyler, %6,5 ile yükseköğretim veya fakülte mezunları, %0,5 ile yüksek lisans mezunları ve %0,09 ile doktora mezunları takip etmektedir.

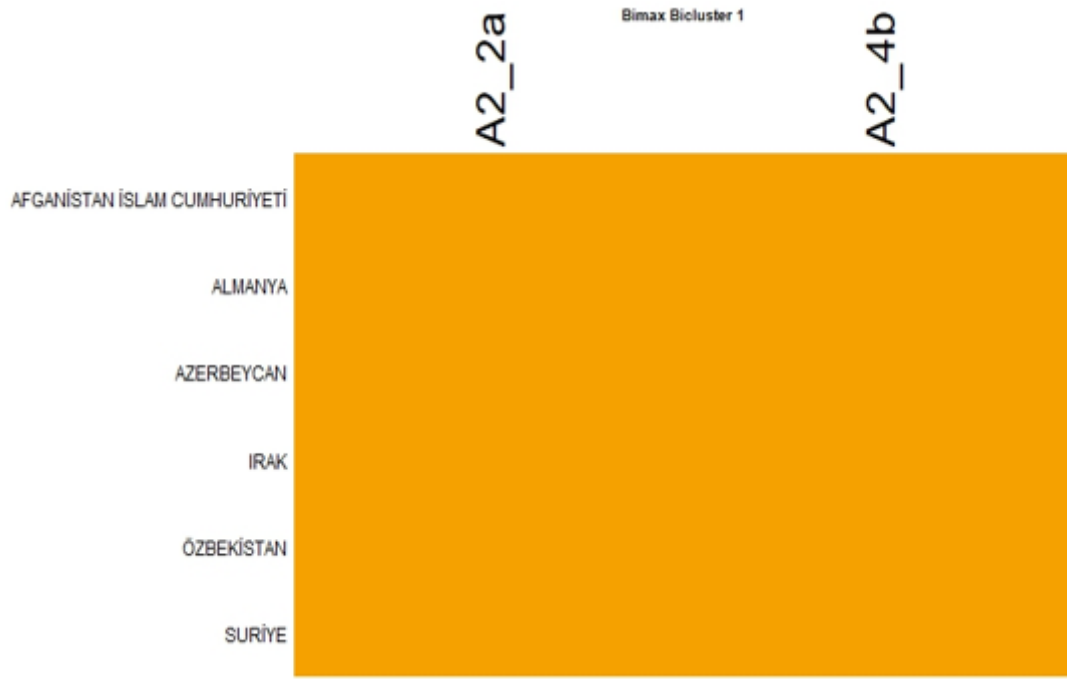
Veriler, suç kaydı bulunan bireylerin önemli bir kısmının düşük eğitim seviyesine sahip olduğunu ortaya koymaktadır. Özellikle ilköğretim mezunları ve okuma-yazma bilmeyenlerin yüksek oranı, eğitimin suç eğilimleri üzerindeki etkisine işaret etmektedir. Eğitim düzeyinin artmasıyla birlikte suç oranlarında belirgin bir düşüş gözlenmekte, yükseköğretim düzeyine sahip bireylerin suç kayıtlarının oldukça düşük seviyelerde kaldığı görülmektedir. Bu durum, eğitimin bireylerin toplumsal entegrasyonu ve suça yönelimlerini azaltmadaki rolünü vurgulamaktadır.

Algoritmada eşik değeri olarak belirlenen 2’ye göre ikili forma dönüştürülmüş; 2’den büyük veya eşit değerlere 1, daha küçük değerlere ise 0 atanmıştır. En az 2 satır ve 2 sütun içeren, en fazla 5 adet ikili küme bulmak için Bimax algoritması ile analiz yapılmıştır.

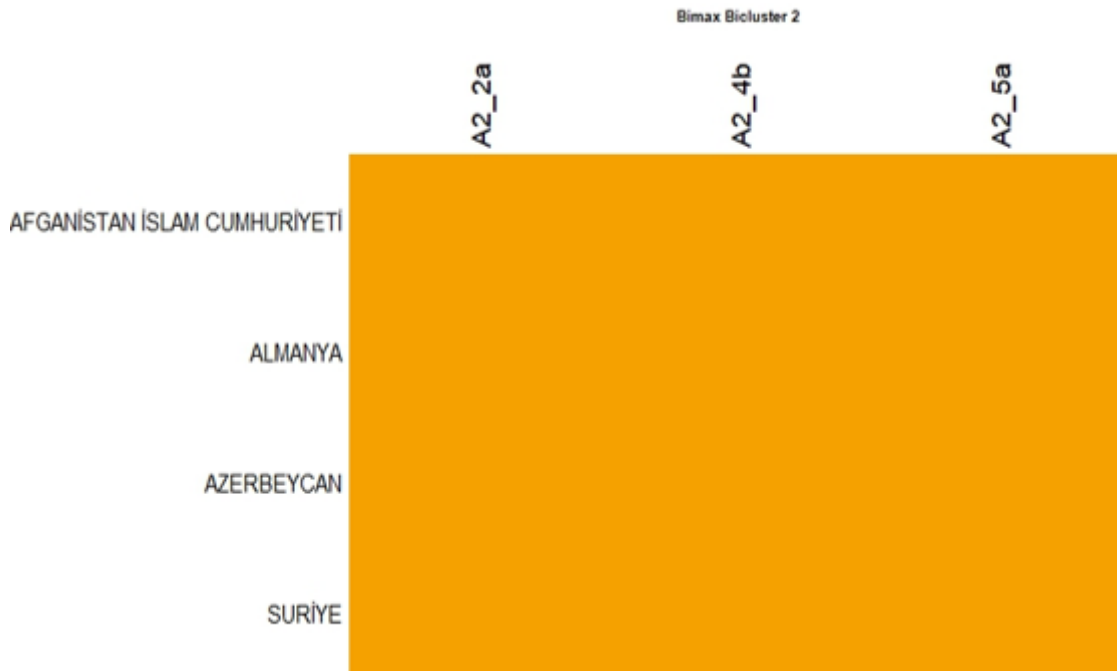
```
biclust_result_Bimax <- biclust(uyruk_binary, method = BCBimax(), minr = 2, minc = 2, number = 5)
biclust_result_Bimax
```

Şekil 3.25. Bimax algoritmasına ait R program kodu

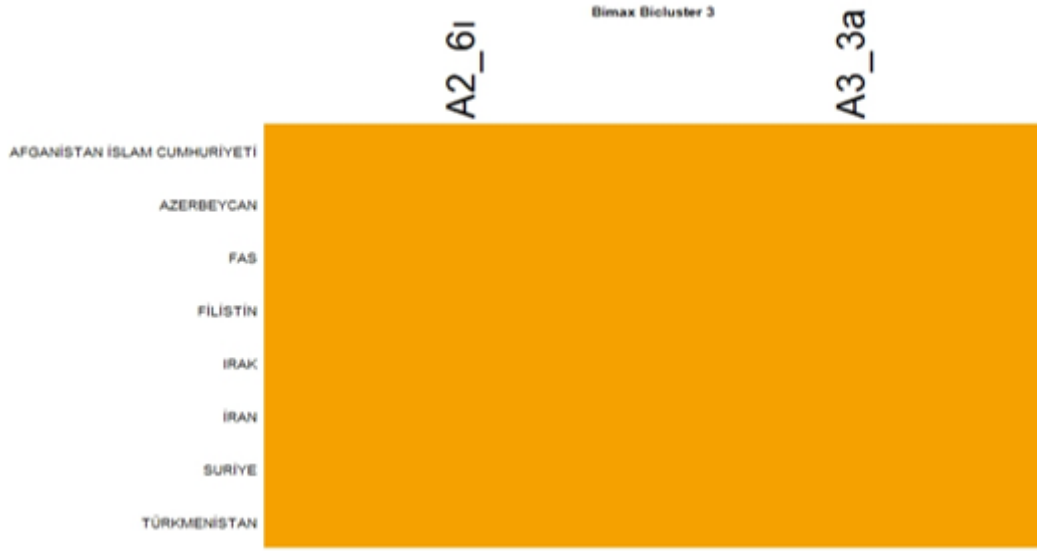
Yukarıda yer alan parametreler sonucunda bulunan 5 adet ikili kümelerin boyutları Çizelge 3.6’da yer almaktadır.



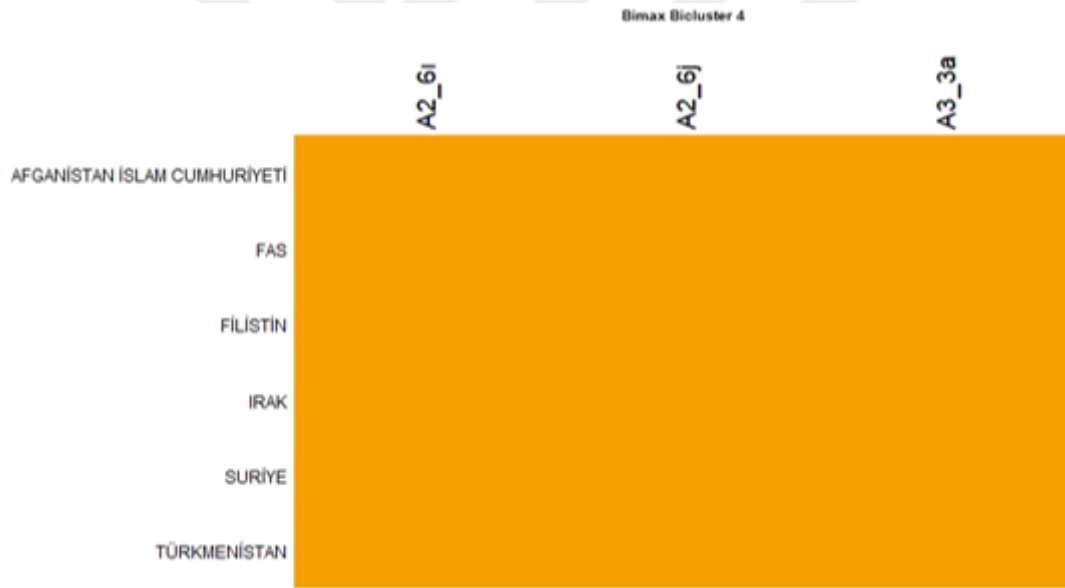
Şekil 3.27. Bimax algoritması ikili küme 1'e ait ısı grafiği



Şekil 3.28. Bimax algoritması ikili küme 2'ye ait ısı grafiği



Şekil 3.29. Bimax algoritması ikili küme 3'e ait ısı grafiği



Şekil 3.30. Bimax algoritması ikili küme 4'e ait ısı grafiği

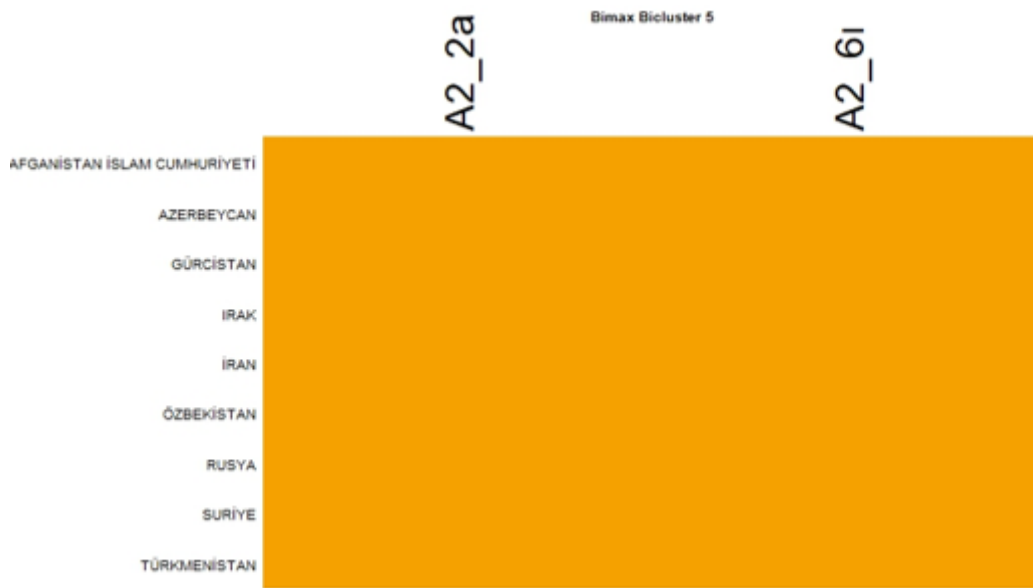
Bimax algoritması ile elde edilen ısı grafikleri incelendiğinde, ikili küme 1'de kasten yaralama ve tehdit suçlarının birlikte kümelendiği, ikili küme 2'de ise bu iki suç türüne ek olarak hakaret suçunun da dâhil olduğu görülmektedir. Bu durum, söz konusu üç suçün çoğu zaman bir arada işleme eğiliminde olduğunu göstermektedir. Benzer bir örüntü, CC algoritması sonuçlarında da gözlemlenmiş olup; bu suçların birlikte kümelmesi, algoritmaların benzer yapısal ilişkileri yakaladığını ve kümelmenin anlamlı bir örüntüye karşılık geldiğini ortaya koymaktadır.

Ayrıca, bu kümeler içerisinde en fazla suç kaydına sahip uyrukların yer alması, Bimax algoritmasının sadece suç türleri arasındaki ilişkileri değil, bu ilişkilerin hangi uyruklar nezdinde yoğunlaştığını da etkili biçimde yakalayabildiğini göstermektedir. Bu durum, ikili kümeleme yönteminin hem yatayda (suç türü) hem dikeyde (uyruk) güçlü ayrımlar yapabildiğini desteklemektedir.

İkili küme 3'te ise farklı bir örüntü dikkat çekmektedir. Nitelikli hırsızlık (A2_6i) ve uyuşturucu veya uyarıcı madde imal ve ticareti (A3_3a) suçları birlikte kümelenmiştir. İkili küme 4'te bu kümeye ek olarak nitelikli yağma (A2_6j) suçunun da dâhil olduğu görülmektedir.

İkili küme 2 ve 4'te oluşan bu genişlemeler kümelerin maksimal ikili kümeleme özelliği kapsamında büyüyerek oluşan yeni bir örüntünün kümelenmeye yansıdığını ortaya koymaktadır.

Bimax algoritmasının maksimal ikili kümeleme yapısı gereği, bu kümelerin her biri en fazla sayıda ortak 1 değeri içeren ve daha büyük bir kümeye genişletilemeyecek yapıdadır. Bu sayede, örtüşen fakat aynı olmayan kümeler oluşturmuş ve hem suç türleri arasında hem de uyruklar nezdinde örtüşme ve ayrışma desenleri gözlemlenmiştir.



Şekil 3.31. Bimax algoritması ikili küme 5'e ait ısı grafiği

Bu ikili kümede, kasten yaralama ve nitelikli hırsızlık (A2_61) suçlarının birlikte kümelendiği gözlemlenmektedir. Bu kümeye ait uyruklar arasında Afganistan, Azerbaycan, Gürcistan, Irak, İran, Özbekistan, Rusya, Suriye ve Türkmenistan vatandaşları yer almaktadır. Bu ülkelerin daha önceki kümelerde de yer almış olması, bazı uyrukların birden fazla suç türünde yüksek temsil düzeyine sahip olduğunu ve örtüşen kümelere dâhil olma eğiliminde olduklarını ortaya koymaktadır.

Bu ısı grafikleri, Bimax algoritmasının hem satır (uyruk) hem de sütun (suç türü) düzeyinde örüntü arayan yaklaşımını başarılı bir şekilde yansıtmaktadır. Özellikle; kasten yaralama, tehdit ve hakaret gibi mağdura yönelen suçların birlikte kümelmesi; nitelikli hırsızlık, uyuşturucu ticareti ve nitelikli yağma gibi organize suçların birlikte kümelmesi ve her bir kümede yoğunlukla yer alan uyrukların suç kayıtlarına göre anlamlı dağılımlar göstermesi bu algoritmanın anlamlı ve örüntü temelli kümeler oluşturduğunu göstermektedir. Aynı zamanda örtüşen kümeler aracılığıyla suç profillerinin çok boyutlu yapısını (bir bireyin birden fazla farklı suç türüne karışabilme ihtimali, aynı suç türünün birden fazla uyrukla ilişkilendirilmesi, bazı uyrukların farklı kümelerde tekrar eden biçimde yer alması (örtüşme), suçların doğası gereği birbirleriyle ilişkili olabilmesi (örneğin tehdit ile kasten yaralama gibi) de açığa çıkarmaktadır.

4. SONUÇ ve ÖNERİLER

Bu çalışmada, Türkiye'deki suç verileri üzerinde farklı ikili kümeleme algoritmaları kullanılarak suç türleri, coğrafi bölgeler (iller) ve uyruklar arasındaki ilişkiler kapsamlı bir şekilde incelenmiştir. Öncelikle, Bimax algoritması ile yapılan analizlerde, özellikle kasten yaralama, tehdit ve hakaret gibi mağdura yönelik suçların belirli uyruklar ile birlikte kümelendiği, ayrıca nitelikli hırsızlık, uyuşturucu ticareti ve nitelikli yağma gibi organize suçların da farklı uyruklar temelinde anlamlı örüntüler oluşturduğu gözlemlenmiştir. Bu algoritmanın örtüşen kümeler aracılığıyla suç profillerinin çok boyutlu yapısını açığa çıkarması, bireylerin birden fazla farklı suç türüne karışabilme ihtimali ve aynı suç türlerinin birden fazla uyrukla ilişkilendirilebilmesi gibi karmaşık ilişkileri başarıyla modellediğini göstermektedir.

xMotif algoritması ile yapılan ikili kümeleme analizinde ise suç türleri ile iller arasındaki ilişkiler, farklı yoğunluk seviyelerine göre kategorize edilmiştir. Bazı kümelerde yüksek yoğunluklu suç türleri (örneğin kasten yaralama, nitelikli hırsızlık, uyuşturucu ticareti) ön plana çıkarken, diğer kümelerde orta ve düşük yoğunluklu suç türleri yoğunlaşmıştır. Bu durum, suçların coğrafi dağılımının homojen olmadığını ve farklı illerde suç tiplerinin değişkenlik gösterdiğini ortaya koymaktadır.

Ancak, veri setinin sayısal değerlerden kesikli (düşük, orta, yüksek) kategorilere dönüştürülmesi sürecinde, yüksek frekanslı suç bölgelerinin bir kısmının kümelene analizde dışlanması gibi sınırlamalar ortaya çıkmıştır. Veri dönüşümü sırasında yaşanan bilgi kaybı, analizlerin bazı ince ayrıntıları kaçırmasına neden olmuş ve yorumlama sürecinde dikkatli olunmasını gerektirmiştir.

CC algoritması sonuçları ise, en fazla işlenen suç türleri ile yoğun suç kaydına sahip illerin birlikte kümelene tutarlı ve anlamlı örüntüler oluşturduğunu göstermiştir. Bu bulgular, algoritmanın veri seti üzerinde etkin şekilde çalıştığını ve tanımlayıcı istatistiklerle uyumlu sonuçlar verdiğini desteklemektedir.

Elde edilen bulgular doğrultusunda, tehdit ve hakaret suçlarının, kasten yaralama suçu ile çoğunlukla birlikte işlendiği gözlemlenmiştir. Bu durum, tehdit ve hakaret suçlarının zamanla kasten yaralamaya dönüşme potansiyeli taşıdığına işaret etmektedir. Dolayısıyla bu

tür suçlar arasında bağlantı kurularak önleyici stratejilerin geliştirilmesi ve kamuoyunda bu konuda farkındalık oluşturulması önemli görülmektedir.

Ayrıca, eğitim düzeyi ve bireylerin maruz kaldığı ekonomik baskıların bazı suç türleri üzerinde belirleyici bir etkisinin olduğu değerlendirilmiştir. Bu nedenle, özellikle düşük gelir gruplarına ve eğitim düzeyi düşük bireylere yönelik sosyal destek programlarının artırılması, suça yönelimi azaltıcı bir unsur olabilir.

Yaş grupları incelendiğinde, belirli yaş aralıklarında bazı suç türlerinin daha sık işlendiği tespit edilmiştir. Bu doğrultuda, özellikle risk grubunda yer alan yaş dilimlerine yönelik önleyici ve eğitici çalışmaların planlanması gerektiği düşünülmektedir.

Veriler, suç türlerinin iller bazında farklılık gösterdiğini ortaya koymuştur. Özellikle büyükşehirlerde nüfus yoğunluğu, göç hareketliliği ve sosyoekonomik eşitsizlikler gibi faktörler, suç oranlarının artmasına neden olabilmektedir.

Çalışmada elde edilen bulgular, suç türlerinin bazı uyruk gruplarında daha yoğun şekilde görülebildiğini ortaya koymuştur. Bu durum, göçmen nüfusun sosyal entegrasyon sürecindeki zorluklar, ekonomik yetersizlikler ve kültürel farklılıklar gibi faktörlerle ilişkili olabilmektedir. Suç oranlarının bazı yabancı uyruklu gruplar arasında yüksek olması, bu topluluklara yönelik sosyal destek mekanizmalarının artırılması gerektiğini göstermektedir. Ayrıca, kültürel uyum programları, dil eğitimi ve istihdam olanaklarının geliştirilmesi gibi önleyici politikaların uygulanması, uzun vadede suç oranlarını düşürmede etkili olabilir.

Sonuç olarak, suçla mücadelede yalnızca hukuki değil, sosyal, ekonomik ve eğitsel faktörlerin de dikkate alındığı bütüncül bir yaklaşım benimsenmesi gerektiği değerlendirilmektedir.

KAYNAKLAR

- Adderley, R. and Musgrove, P.B, (2001). Police crime recording and investigation systems, a user's view. *Policing: An International Journal of Police Strategies and Management*, 24(1), 100-114.
- Adalet Bakanlığı Adli Sicil ve İstatistik Genel Müdürlüğü, (2023). *Adli İstatistikler 2023*.
- Aitkin, M., Bock, R. D. (2015). Variational algorithms for biclustering models. *Computational Statistics Data Analysis*, 89, 12–24.
- Ateş, H., Bostancı, E., Mşg, S. (2020). Suç verisi analizinde büyük veri, veri madenciliği ve yapay zekâ uygulamaları. *Polis Bilimleri Dergisi*, 22(2), 123–142.
- Bappee, M. H., Soares Junior, D. S., Matwin, S. (2018). Crime prediction based on spatial features extracted from OpenStreetMap. In *2018 IEEE International Conference on Big Data (Big Data)*, 1637–1642, IEEE.
- Brown, D. E. (1998). The Regional Crime Analysis Program (RECAP): A framework for mining data to catch criminals. *IEEE International Conference on Systems, Man, and Cybernetics*, 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., Pardalos, P. M. (2008). Biclustering in data mining. *Computers and Operations Research*, 35(9), 2964-2987.
- Cheng, Y., Church, G. M. (2000). Biclustering of expression data. In *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*, 8, 93–103.
- Chen, H., Chung, W., Qin, Y., Chau, M., Xu, J.J., Wang, G., Zheng, R. ve Atabakhsh, H., (2003). Crime data mining: An overview and case studies. *ACM International Conference Proceeding Series*, Vol. 130.
- Cömertler, N. ve Kar, M. (2010). Türkiye’de suç oranının sosyo-ekonomik belirleyicileri. *Ankara Üniversitesi Sosyal Bilimler Fakültesi Dergisi*, 67, 62-102.
- Çalışkan, M., Yıldız, M., Doğan, S., Aktaş, E. (2021). Suç verileri üzerinden birliktelik kuralları ile bilgi çıkarımı: Maryland örneği. *Journal of Intelligent Systems: Theory and Applications (JISTA)*, 4(2), 45–56.
- Emniyet Genel Müdürlüğü Narkotik Suçlarla Mücadele Daire Başkanlığı. (2023). *2023 Türkiye Uyuşturucu Raporu*. URL: <https://www.narkotik.pol.tr/kurumlar/narkotik.pol.tr/TUB%C4%B0M/Ulusal%20Yay%C4%B1mlar/NARKOLOG-2023-PROFIL-ANALIZI.pdf>

- Fernandes, R., Freitas, A. A. (2008). Applying biclustering to perform collaborative filtering. *Proceedings of the 2008 ACM Symposium on Applied Computing*, 1381–1385.
- Giray, S. (2016). İki Aşamalı Kümeleme Analizi ile Hükümlü Verilerinin İncelenmesi. *Ekonometri ve İstatistik*, 25, 1-31.
- Henriques, R., Antunes, C., Madeira, S. C. (2015). A structured view on pattern mining-based biclustering. *Pattern Recognition*, 48(12), 3941–3958.
- Kadar, C., Pletikosa, I. (2018). Predicting crime using spatial features derived from social media and transportation data. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 1–4.
- Kasim, A., Shkedy, Z., Kaiser, S., Hochreiter, S., Talloen, W. (2016). *Applied biclustering methods for big and high-dimensional data using R*.
- Kocatürk A., Altunkaynak B., Örkücü H. H., Arslan R. (2017). Suç Verilerinin Analizinde Kullanılan Veri Madenciliği Yöntemleri ve Kaçakçılık Suçlarıyla İlgili Bir Uygulama. In *2nd International Mediterranean Science and Engineering Congress (IMSEC)*.
- Kocatürk A., Altunkaynak B. (2018). Gen Açıklama Verileri İçin İkili Kümeleme Algoritmalarının Karşılaştırılması ve Uygulanması. *Türkiye Klinikleri Journal of Biostatistics*, 10(2), 137-152.
- Kocatürk A., Altunkaynak B., Örkücü H. H. (2018). Türkiye’de Sektörlerin Ölümlü Kaza Türlerine Göre Kümelenmesi: İkili Kümeleme Yöntemi. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (20), 871-882.
- Lazzeroni, L., Owen, A. (2002). Plaid models for gene expression data. *Statistica Sinica*, 12, 61-86.
- Liu, B., Yu, C. W., Wang, D. Z., Cheung, R. C. C., Yan, H. (2014). Design exploration of geometric biclustering for microarray data analysis in data mining. *IEEE Transactions on Dependable and Secure Computing*, 11(5), 440–452.
- Ma, L., Chen, Y. ve Huang, H., (2010). AK-Modes: A weighted clustering algorithm for finding similar case subsets. *IEEE*, 218-223.
- Madeira, S. C., Oliveira, A. L. (2004). Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1(1), 24–45.
- Marin, J., Guerreros, L., Calderon, D. (2024). Crime trend analysis using big data and clustering methods. *International Journal of Data Science and Analytics*, 12(1), 33-47.
- Martella, F., Vichi, M. (2010). Hierarchical mixture models for biclustering in microarray data. *Statistical Modelling*, 11(6), 489–512.
- Mirkin, B. (1998). Mathematical classification and clustering: From how to what and why. *Classification, Data Analysis, and Data Highways*, 172-181.

- Murali, T. M. and 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. *IEEE/WIC/ACM, International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops*, 41-44.
- Oğuzlar, A., (2005). Kümeleme analizinde yeni bir yaklaşım: Kendini düzenleyen haritalar (Kohonen Ağları). *İktisadi ve İdari Bilimler Dergisi*, 19(2), 1-12.
- Orakcı, M., Ciylan, B., Kök, İ. ve Sevri, M., (2016). Suç Analizinde Veri Madenciliği Teknikleri ve Makine Öğrenmesi Algoritmalarının Kullanılması, 18. *Akademik Bilişim Konferansı*, Adnan Menderes Üniversitesi, Aydın.
- Padilha, V. A., Campello, R. J. G. B. (2017). A systematic comparative evaluation of biclustering techniques. *BMC Bioinformatics*, 18, 55.
- Polat, C., Eren, H. ve Erbakıcı, F., (2013). Hırsızlık Suçunu Etkileyen Faktörlerin Değerlendirilmesi ve Geleceğe Yönelik Yaklaşımlar. *Güvenlik Bilimleri Dergisi*, 2(1), 1-33.
- Prelić, A., Bleuler, S., Zimmermann, P., Wille, A., Bühlmann, P., Gruissem, W., Zitzler, E. (2006). A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics*, 22(9), 1122–1129.
- Puolamäki, K., Hanhijärvi, S., Garriga, G. C. (2007). An approximation ratio for biclustering. In *Proceedings of the 24th International Conference on Machine Learning (ICML 2007)*, 737–744.
- Reale, K., Beauregard, E. ve Martineau, M., (2017). Sadism in sexual homicide offenders: identifying distinct groups. *Journal of Criminal Psychology*, 7(2), 120-133.
- Sariyüce, A. E., Kaya, K. (2015). Detection of crime regions with biclustering approach and comparison of methods. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 1011–1016.
- 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.
- Sevüktekin, M., Oğuzlar, A., Aydın, B. ve Nargeleçekenler, M., (2007). Karar Ağacı Yardımıyla Suçluların Özelliklerinin Belirlenmesi. *Öneri Dergisi*, 7(27), 291-298.
- Stalidis, G., Semertzidis, T., Daras, P. (2018). Criminal incidents prediction using deep learning. In *Proceedings of the 15th International Conference on Content-Based Multimedia Indexing*, 1–6.
- Stec, K., Klabjan, D. (2018). Predicting crime using deep learning with spatio-temporal data. *arXiv preprint arXiv:1802.05957*.
- Tanay, A., Sharan, R., Shamir, R. (2002). Discovering statistically significant biclusters in gene expression data. *Bioinformatics*, 18(1), 136–144.

Tüzüntürk, S., (2009). Çok boyutlu ölçekleme analizi: suç istatistikleri üzerine bir uygulama. *Uludağ Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Dergisi*, 28(2), 71-91.

Zhao, Y., Zaki, M. J. (2011). A biclustering algorithm for extracting bit-patterns from binary datasets. *Bioinformatics*, 27(19), 2738–2745.





Gazili olmak ayrıcalıktır