

ÇOK DEĞİŞKENLİ HARİTALAMA İÇİN KÜMELEME YÖNTEMLERİNİN KULLANILMASI

Hüseyin Zahit SELVİ (ORCID:0000-0001-7486-0992)^{1*}
Burak ÇAĞLAR (ORCID:0000-0002-4490-1447)²

¹Harita Mühendisliği Bölümü, Mühendislik ve Mimarlık Fakültesi, Necmettin Erbakan Üniversitesi, Konya, Türkiye
²Çorum İl Özel İdaresi, Yol ve Ulaşım Hizmetleri Müdürlüğü, Çorum, Türkiye

Geliş / Received: 25.11.2016

Düzeltilmelerin gelişi / Received in revised form: 07.01.2017

Kabul / Accepted: 19.01.2017

ÖZ

Çok değişkenli haritalama mekânsal objelere ait birden çok özelliğin harita kullanılarak görsel sunumudur. Çeşitli veri hazırlama ve istatistiksel sınıflandırma teknikleri kullanılarak mekânsal objelere ait birden çok özellik görsel olarak incelenebilir ve kartografik işaretlerle gösterilebilir. Bu kapsamda kümeleme analizi yöntemleri de çok değişkenli haritalama için kullanılabilir. Bu çalışmada kümeleme analiz yöntemlerinden k-ortalama yöntemi, k-temsilci yöntemi ve Birleştirici Hiyerarşik Kümeleme yöntemi ele alınmıştır. Bu yöntemlerle Türkiye'deki üç ayrı yıla ait trafik kaza verileri kullanılarak oluşturulan sınıflar ve üretilen çok değişkenli haritalar kullanılarak bu yöntemlerin karşılaştırılması yapılmış, bu yöntemlerle üretilen haritaların risk yönetimi ve planlamada kullanılabilirliği üzerinde durulmuştur.

Anahtar Kelimeler: Çok değişkenli haritalama, veri madenciliği, kümeleme analizi

USING CLUSTER ANALYSIS METHODS FOR MULTIVARIATE MAPPING

ABSTRACT

Multivariate mapping is the visual exploration of spatial objects with multiple attributes using a map. More than one attribute can be visually explored and symbolized using numerous statistical classification systems or data reduction techniques. In this sense, clustering analysis methods can be used for multivariate mapping. In this study, among clustering analysis methods, k-means method, k-medoids method and Agglomerative Hierarchical Clustering method were selected. For this purpose, multivariate maps created from traffic accident data of two different years in Turkey were used. The methods were compared using the maps produced with these methods and effectiveness of these maps in risk management and planning were discussed.

Keywords: Multivariate mapping, data mining, cluster analysis

1. GİRİŞ

Çok değişkenli haritalama mekânsal objelerin birden fazla özelliğinin harita üzerinde gösterilmesidir. Birden fazla özelliğin eş zamanlı gösterimi mekânsal objelerin farklı özellikleri dikkate alınarak karşılaştırılmalarına olanak sağlar. Çok değişkenli haritalama için kartografik gösterimi sağlamak amacıyla farklı bilgisayar destekli yöntemler geliştirilmiştir [1].

*Corresponding author / Sorumlu yazar. Tel.: +90 332 325 2024; e-mail / e-posta: hzselvi@konya.edu.tr

Çok değişkenli haritalamada her bir özellik için ayrı bir harita mı yapılacağına yoksa birçok özelliğin aynı haritada mı gösterileceğine karar vermek gerekir [2]. Her bir özellik için ayrı harita yapılması, çeşitli özelliklere sahip iki objenin karşılaştırılmasını zorlaştırabilir. Bu nedenle birçok özelliğin aynı haritada gösterildiği yöntemler daha çok tercih edilmektedir. Bu kapsamda iki renkli koroplek haritanın üst üste çakıştırılmasıyla oluşturulan Üç Değişkenli Koroplek Harita (Trivariate Choropleth Map) yöntemi [3-6], haritadaki her bir özellik için belirli bir renk veya şekildeki işaretin kullanıldığı Çok Değişkenli Nokta Haritalar (Multivariate Dot Maps) yöntemi [7], çok değişkenli verinin noktasal işaretlerle gösterilmesi uygun olan durumlarda kullanılan Çok Değişkenli Noktasal İşaret (Multivariate Point Symbol) yöntemleri [8-15], çok değişkenli verinin çeşitli işaretlerin birleştirilmesiyle gösterildiği Farklı İşaretlerin Birleştirilmesi (Combining Different Type of Symbols) yöntemi [16] ve değişik özelliklerin birleşik işaretlerden ayırt edilmesi (Seperable Versus Integral Symbols) yöntemi [17] sayılabilir.

Yukarıda sayılan yöntemlerin genel olarak çok büyük veri gruplarında uygulanmasında ve haritaların okunabilirliğini sağlamakta zorluklar yaşanmaktadır [18, 19]. Bu yöntemlerden farklı olarak aynı haritada birçok özelliği gösterebilmek için veri madenciliğindeki kümeleme yöntemi esaslı sınıflandırma yöntemleri de kullanılabilir [2]. Kümeleme yöntemlerinin kullanılmasıyla, büyük veri gruplarında da mekânsal verilerin birden fazla özelliği dikkate alınarak ve bu özelliklerin birbirleriyle olan ilişkileri ortaya konarak farklı mekânsal verilerin benzer yönleri ortaya çıkarılabilir. Bu sayede risk analizi, planlama vb. uygulamalar için önemli katkı sağlayacak mekânsal analizler yapılabilir. Murray ve ark. [20] ve Grubestic ve ark. [21] yayınlarında seçilen kümeleme yöntemleriyle mekânsal analizler yapılmış ve kümeleme yöntemlerinin oluşturulan yapay verilerle hesaplama gücü ve kümeleme başarısı açısından karşılaştırılması yapılmıştır. Guo ve ark. [18] yayınında ise kümeleme yöntemlerine benzer şekilde iteratif bir yol izleyen Coğrafi Bilgi Çıkarımı (geographic knowledge discovery) yöntemini kullanarak çok değişkenli haritalar üretmiştir. Brimicombe [22] ve Anderson [23] yayınlarında trafik kazalarının yoğun olduğu noktalar k-ortalama kümeleme yöntemiyle belirlenmiş, Lu ve ark. [24] yayınında sürekli artan verilerin kümelemesi üzerinde durmuşlar ve yollarda hareket eden taşıtlardan aldıkları sürekli GPS sinyallerini kullanarak yolları yoğunluklarına göre kümelemişlerdir. Weng ve ark. [25] trafik kazalarının oluş sürelerini kümelemeye dayalı olarak incelemiş, Guo ve Fang [26] sürücülerden kaynaklı riskleri k-ortalama yöntemiyle kümeleme analizi yaparak araştırmış, Yalcin ve Duzgun [27] iki tekerlekli araçların trafik kazalarının mekânsal analizlerini gerçekleştirmiş, Feng ve ark.[28] otobüs kazalarının oluşmasındaki risk faktörlerini k-ortalama yöntemiyle kümeleyerek irdelemiş, Erdogan [29] yollardaki ölüm oranını coğrafi ağırlıklı regresyon analiziyle araştırmış, Martinussen ve ark. [30] sürücü davranışları ve becerileri arasındaki ilişkiyi k-ortalama yöntemiyle ortaya koymuşlardır. Bu çalışmalarda genel olarak bir veya iki değişkene bağlı olarak analizler yapılmış, yine seçilen bir kümeleme yöntemiyle sonuçlar ortaya konulmuş ve tematik haritalar üretilmiştir. Bu çalışmada yukarıdaki çalışmalardan farklı olarak hiyerarşik ve hiyerarşik olmayan kümeleme yöntemlerinin çok değişkenli haritalama için kullanılabilirliği ve bu yöntemlerle üretilen çok değişkenli haritaların gerçek verilerle karşılaştırılarak risk yönetimi ve planlama açısından doğruluğu araştırılmakta, hangi kümeleme yöntemlerinin çok değişkenli haritalama açısından tercih edilmesinin daha doğru olduğuna yönelik öneriler sunulmaktadır. Bu kapsamda Türkiye'deki trafik kazaları şehir bazında motorlu araç sayısı, ölüm ve yaralanmalı kaza sayısı, ölü sayısı ve yaralı sayısı parametreleri dikkate alınarak incelenmiştir. Bu 4 değişken dikkate alınarak, hiyerarşik ve hiyerarşik olmayan kümeleme yöntemleri ve 3 farklı yıla ait veriler kullanılarak, 3 farklı yıl için çok değişkenli haritalar üretilmiştir. Üç farklı yıl için üretilen sonuç haritalar karşılaştırılarak bu yöntemlerin hangilerinin çok değişkenli haritalama açısından tercih edilebileceği ortaya konulmuştur. Çalışmanın ikinci bölümünde kullanılan kümeleme yöntemleri, üçüncü bölümde yapılan uygulamaya ayrıntılı olarak anlatılacak, sonuç bölümünde elde edilen bulgular ve öneriler paylaşılmıştır.

2. MATERYAL VE METOT

2.1. Kümeleme Analizi

Kümeleme analizi, bir veri kümesindeki bilgileri belirli yakınlık kriterlerine göre gruplara ayırma işlemidir. Aynı küme içindeki elemanların benzerliği fazla, kümeler arası benzerlik ise az olmalıdır [31]. Sınıflandırma işleminde, sınıflar önceden bellidir. Kümeleme yönteminde ise sınıflar önceden belli değildir. Verilerin hangi kümelere, kaç değişik gruba ayrılacağı eldeki verilerin birbirlerine olan benzerliğine göre belirlenir.

Kümeleme yöntemleri çeşitli kaynaklarda farklı biçimlerde sınıflandırılmaktadır. Genel olarak kümeleme yöntemleri hiyerarşik olmayan ve hiyerarşik yöntemler olarak sınıflandırılabilir [2].

- **Hiyerarşik Olmayan Yöntemler:** Hiyerarşik olmayan yöntemlerde n adet nokta önceden verilen k küme sayısına ($k < n$) göre kümelere ayrılır. Bu yöntem veriyi her grupta en az bir obje olacak ve her obje en az bir gruba dahil olacak şekilde gruplara ayırır [32].

ÇOK DEĞİŞKENLİ HARİTALAMA İÇİN KÜMELEME YÖNTEMLERİNİN KULLANILMASI

• **Hiyerarşik Yöntemler:** Hiyerarşik kümeleme metodu veri nesnelere ağaç yapısı içinde gruplamaya çalışır. Hiyerarşik kümeleme metodları hiyerarşik ayrışmanın aşağıdan yukarı (bottom-up) ya da yukarıdan aşağıya (top-down) olmasına bağlı olarak birleştirici (agglomerative) ya da ayırıcı (divisive) olmak üzere sınıflandırılırlar [33].

Bu çalışmada hiyerarşik olmayan yöntemlerden k-Ortalama Yöntemi (The k-Means Method) ve k-Temsilci Yöntemi (k-Medoids Method), hiyerarşik yöntemlerden Birleştirici ve Ayırıcı Hiyerarşik Kümeleme Yöntemi (Agglomerative and Divisive Hierarchical Clustering) ele alınacaktır.

2.1.1. k-Ortalama Yöntemi

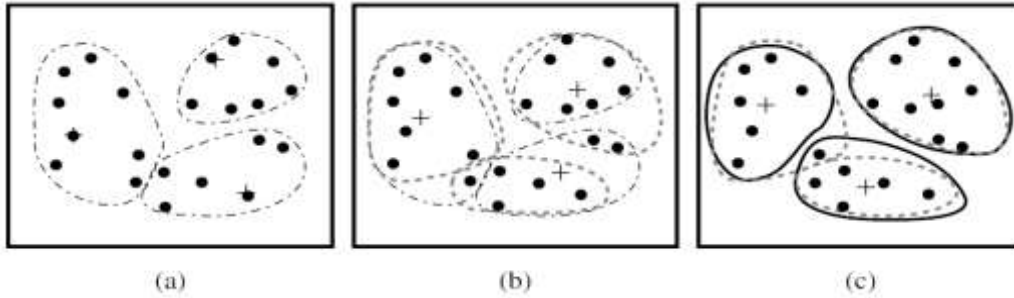
İlk olarak 1967 yılında Mac Quenn tarafından ortaya atılan bu algoritma sürekli olarak kümelerin yenilendiği ve en uygun çözüme ulaşana kadar devam eden döngüsel bir algoritmadır. k-Ortalama algoritmasının genel mantığı n adet veri nesnesinden oluşan bir veri kümesini, araştırmacının ön bilgisine ve tecrübesine dayanarak belirlenen k adet kümeye bölümlenektir. Amaç, gerçekleştirilen bölümlenme işlemi sonunda elde edilen kümelerin küme içi benzerliklerini maksimum ve farklı kümeler arası benzerliklerin minimum olmasını sağlamaktır. Kümenin benzerliği kümedeki objelerin ortalama değeriyle (mean value) ölçülür (Şekil 1).

k-ortalama yönteminin işlem adımları aşağıdaki gibi özetlenebilir [32]:

k: küme sayısı

$D = \{t_1, t_2, \dots, t_n\}$: n sayıda objesi olan bir veri seti olmak üzere;

1. D veri setinden keyfi olarak belirlenen kümeler için başlangıç ortalama değerleri m_1, m_2, \dots, m_k belirlenir.
2. Her bir t_i en yakın olduğu m_i 'nin kümesine atanır.
3. Kümelere ait m_1, m_2, \dots, m_k değerleri yeniden hesaplanır.
4. Küme elemanlarında herhangi bir değişiklik yoksa durulur.
5. Değişiklik varsa ilk adımdan itibaren tekrar edilir.



Şekil 1. (a, b, c) k-Ortalama Algoritması ile kümeleme aşamaları [33]

2.1.2. k-Temsilci Yöntemi

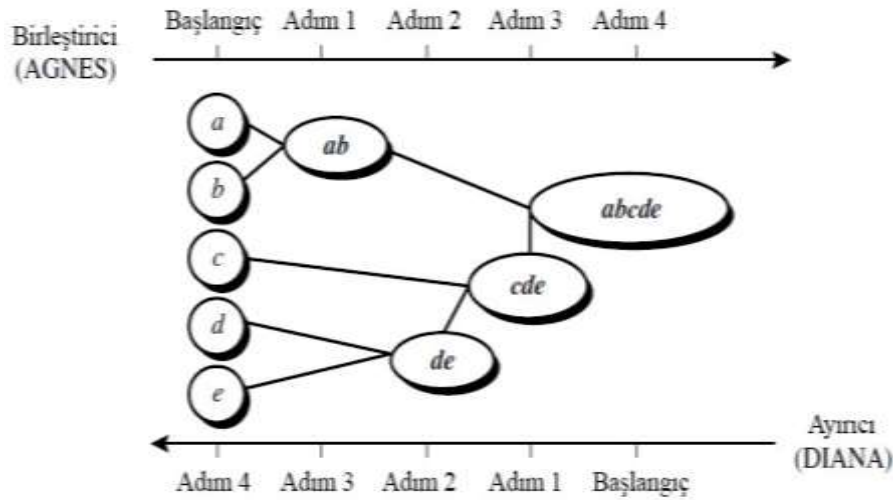
1990 yılında Kauffman ve Rousseuw tarafından geliştirilen bu algoritmada her bir kümenin ortalama değeri yerine her bir kümedeki bir obje referans olarak alınmaktadır. Referans alınan bu objeye temsilci (medoid) denilmektedir [32]. k adet kümeyi bulmak için seçilen temsilcilerin etrafına ana kümedeki tüm elemanları toplayarak ve her defasında bu temsilcileri değiştirilerek kümeleme işlemi tamamlanmaktadır. Temsilci seçiminden kasıt kümenin merkezine yakın mesafede bulunan noktanın belirlenmesidir. k adet küme için seçilen k adet temsilci belirlendikten sonra, veri tabanındaki temsilci olmayan diğer noktalar (veriler) kendilerine en çok benzeyen temsilcinin etrafında toplanır [34].

k-temsilci algoritmasının işlem basamakları aşağıdaki gibi özetlenebilir[35]:

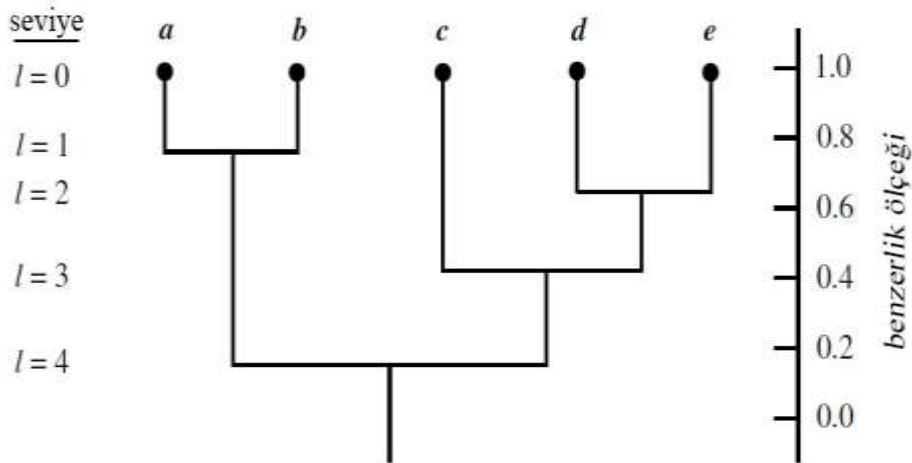
1. k küme sayısının belirlenir.
2. Başlangıç temsilcileri olarak k nesnelere seçilir.
3. Kalan nesnelere en yakın temsilcisine sahip kümeyle atanır.
4. Amaç fonksiyonu hesaplanır (bütün nesnelere en yakın temsilciye olan uzaklıklarının toplamı).
5. Rastgele temsilci olmayan y noktası seçilir.
6. Eğer x ile y'nin yer değiştirmesi amaç fonksiyonunu minimize edecekse bu iki noktanın (x ile y) yerleri değiştirilir.
7. Değişiklik olmayana kadar 3 ile 6. adım arası işlemler tekrarlanır.

2.1.3. Birleştirici ve Ayırıcı Hiyerarşik Kümeleme Yöntemi

Birleştirici Hiyerarşik Kümeleme Yöntemi AGNES (AGglomerative NESTing) aşağıdan yukarıya doğru çalışan bir strateji izler. Başlangıçta her nesne ayrı bir küme olarak kabul edilir. Algoritmanın bir sonraki her adımında bu kümelerden benzerlik gösterenler tek bir küme oluncaya kadar ya da istenen özellikleri sağlayana kadar birleştirilirler. Hiyerarşik kümeleme yöntemlerinin çoğu bu kategoride yer almaktadır. Diğer yandan, Ayırıcı Hiyerarşik Kümeleme Yöntemi DIANA (DIvisive ANAlysis) yukarıdan aşağı çalışan bir strateji izler. Başlangıçta veri nesnelerinin tümü bir küme olarak kabul edilir. Algoritmanın bir sonraki her adımında kendi aralarında benzerliklerin en yüksek olan nesnelere bir araya getirilerek büyük küme daha küçük kümelere bölünür. Bu kümeleme işlemi, her nesne kendi başına bir küme oluşturana kadar ya da istenen herhangi bir şart sağlanana kadar devam eder [33] (Şekil 2). Hiyerarşik kümelemenin sürecini göstermek üzere dendrogram olarak adlandırılan bir ağaç yapısı kullanılır. Dendrogram nesnelerin adım adım nasıl gruplandığını gösterir (Şekil 3).



Şekil 2. {a,b,c,d,e} verisi üzerinde Birleştirici ve Ayırıcı Hiyerarşik Kümeleme



Şekil 3. {a,b,c,d,e} verisinin Hiyerarşik Kümelenmesi için dendrogram gösterim

Hiyerarşik kümeleme yöntemi, basit olarak düşünülse de birleştirme veya bölme noktalarının seçimi konusunda zorluklarla karşılaşılır. Bu noktaların seçimi, sonraki adımlar, bir nesne grubunun bölünmesiyle ya da birleştirilmesiyle oluşturulan yeni kümeler üzerinden yürütüleceği için çok önemlidir. Ne önceki işlemleri ne de kümeler arasındaki nesnelere değiştirmek mümkündür. Bu yüzden birleştirme ya da bölme kararlarının, bazı adımlarda verilmemesi düşük kaliteli kümelerin oluşmasına neden olabilmektedir.

ÇOK DEĞİŞKENLİ HARİTALAMA İÇİN KÜMELEME YÖNTEMLERİNİN KULLANILMASI

3. BULGULAR VE TARTIŞMA

3.1. Uygulama

Trafik kazaları sonucu oluşan ölümler, yaralanmalar ve maddi hasarlar Türkiye'nin en önemli sorunlarından. Son 5 yıla ait veriler incelendiğinde her yıl 1.000.000'dan fazla trafik kazası meydana geldiği, bunların ortalama 145.000'nin ölümlü ve yaralanmalı kaza olduğu, yaklaşık 1.060.000'nin de maddi hasarlı kaza olduğu görülmektedir. Bu kazalarda ortalama yılda 4.000 kişi hayatını kaybetmekte, yaklaşık 250.000 kişi yaralanmaktadır. Türkiye'de trafik güvenliğinin sağlanabilmesi amacıyla alınması gereken önlemler ve yapılacak yatırımların belirlenmesi için birden çok mevcut trafik kaza verisinden yararlanarak, hangi illerdeki kaza verilerinin benzerlik taşıdığı belirlenmesi oldukça önemlidir. Bu amaçla bu çalışmada Türkiye İstatistik Kurumu (TUİK) tarafından hazırlanan 2011, 2012 ve 2013 yıllarına ait il bazlı motorlu kara taşıtı sayısı, ölümlü ve yaralanmalı trafik kaza sayıları, ölü ve yaralı sayıları verileri kullanılarak (4 farklı değer) 3 farklı yöntemle kümeleme analizi yapılmış ve kümeleme analizi sonuçlarına göre çok değişkenli haritalar üretilmiştir. 3 farklı yöntemle her üç yıla ait üretilen haritalar karşılaştırılarak çok değişkenli haritalama ve kümeleme başarısı açısından hangi yöntemin daha uygun olduğu değerlendirilmiştir.

Kümeleme Analiz yöntemlerinin uygulanmasında IBM şirketinin geliştirmiş olduğu IBM SPSS (Statistical Package for the Social Sciences) İstatistik yazılımı [36] ve Ralf Klinkenberg, Ingo Mierswa ve Simon Fischer tarafından Dortmund Teknoloji Üniversitesi Yapay Zeka biriminde geliştirilen RapidMiner [37] yazılımı kullanılmıştır. ESRI grubunun geliştirmiş olduğu ArcGIS [38] yazılımıyla da çok değişkenli haritalar üretilmiştir.

3.2. Verilerin Hazırlanması

Kümeleme analizleri yapılmadan önce ilk olarak veriler arasında birim farklılıkları varsa veya herhangi bir veri grubundaki sayısal değerler diğer veri grubuna göre çok büyükse kümeleme sonuçlarının olumsuz etkilenmemesi için verilerin standardize edilmesi gerekir. Çalışmada kullanılan 4 özelliğe ait veriler aynı birimde olduğundan verilerin birim birlikteliği için standardize edilmesine gerek duyulmamıştır. Ancak çalışmada kullanılan şehir bazında motorlu araç sayısı verileri diğer verilerden oldukça büyüktür. Bu durumun kümeleme sonuçlarını olumsuz etkilememesi için şehir bazında motorlu araç sayısı verileri, Türkiye'deki toplam motorlu araç sayısına bölünerek standardize edilmiştir.

Yine kümeleme analizi yapılmadan önce kullanılacak verilerin birbiriyle ilişkili (korelasyonlu) olup olmadığının test edilmesi gerekir. Çünkü her türlü veri birbiriyle ilişkili olsun veya olmasın kümeleme algoritmalarında analiz edilerek sonuç alınabilir. Gerçekte birbiriyle ilişkili olmayan verilerin kümeleneceği sonucu ise istenmeyen sonuçlar ortaya çıkabilir. Çalışmadaki verilerin birbiriyle ilişkili olup olmadığını belirlemek için verilerin birbirleriyle korelasyon katsayıları 1 numaralı eşitliğe göre hesaplanmıştır.

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

Eşitlikte r iki değişken arasındaki korelasyon katsayısını, X_i ve Y_i değişkenleri, \bar{X} ve \bar{Y} değişkenlerin ortalama değerlerini göstermektedir. Hesaplamalar sonucu 2011, 2012 ve 2013 yılları için aşağıdaki değerler bulunmuştur:

2011 yılına ait veri seti üzerinde yapılan uygulamada;	
Motorlu taşıt sayısı-Ölümlü yaralanmalı kaza sayısı korelasyonu	$r1=0,96$
Motorlu taşıt sayısı-Ölü sayısı korelasyonu	$r2=0,83$
Motorlu taşıt sayısı-Yaralı sayısı korelasyonu	$r3=0,94$
2012 yılına ait veri seti üzerinde yapılan uygulamada;	
Motorlu taşıt sayısı-Ölümlü yaralanmalı kaza sayısı korelasyonu	$r1=0,95$
Motorlu taşıt sayısı-Ölü sayısı korelasyonu	$r2=0,88$
Motorlu taşıt sayısı-Yaralı sayısı korelasyonu	$r3=0,93$
2013 yılına ait veri seti üzerinde yapılan uygulamada;	
Motorlu taşıt sayısı-Ölümlü yaralanmalı kaza sayısı korelasyonu	$r1=0,94$
Motorlu taşıt sayısı-Ölü sayısı korelasyonu	$r2=0,89$
Motorlu taşıt sayısı-Yaralı sayısı korelasyonu	$r3=0,93$

Romesburg [39]'un kümeleme çalışmalarında kullanılacak verilerin ilişkili olması için korelasyon katsayısının 0,80 ve üzeri olması gerektiği görüşü dikkate alındığında çalışmada kullanılan veri setindeki değişkenlerin

birbiriyle ilişkili olduğu görülmektedir. Bu şekilde kümele öncesi yapılması gereken iki aşama tamamlanmış ve kümeleme aşamasına geçilmiştir.

3.3. Küme Sayılarının Belirlenmesi için Parametre Analizi

3.3.1 k-Ortalama ve k-Temsilci Yöntemi için Parametre Analizi

k-Ortalama ve k-Temsilci yönteminde küme sayısı kullanıcı tarafından belirlenir. Ancak veri madenciliğinde küme sayısı önemlidir. Kullanıcı ya deneme yanılma yöntemiyle ya da her kümeleme sonrası bir takım testler yaparak en uygun küme sayısını belirleyebilmektedir. Bu çalışmada küme sayısının belirlenmesinde Dunn Geçerlilik indeksi ve Davies-Bouldin Geçerlilik İndeksi testlerinden yararlanılmıştır.

Dunn Geçerlilik indeksinin temel varsayımı, kümelerin yoğun ve iyi dağılmış olmalarıdır. Dunn Geçerlilik indeksi D katsayısı ile gösterilir, D katsayısı büyüdükçe, küme kalitesi ve sayısı artmaktadır.

$$D = \min_{1 \leq i \leq n} \left\{ \begin{matrix} \min_{1 \leq j \leq n} d(c_i, c_j) \\ \max_{i=j} (d'(c_k)) \\ 1 \leq k \leq n \end{matrix} \right\} \quad (2)$$

Denklemden $d(c_i, c_j)$ c_i ve c_j kümeleri arasındaki mesafeyi, $\max(d'(c_k))$ k kümesinin noktaları arasındaki en uzak mesafeyi ve n küme sayısını belirtmektedir. Algoritmanın amacı kümeler arası mesafeyi en küçüğe çekerken, küme içi mesafeyi maksimumda tutmaktır. Bu sayede elde edilecek D değeri yükseldikçe en uygun küme sayısına yaklaşılmaktadır [34].

Davies-Bouldin Geçerlilik indeksi DB katsayısıyla gösterilmektedir.

$$DB = \frac{1}{n} \sum_{i=1}^n \max \left\{ \frac{S_n(Q_i) + S_n(Q_j)}{S_n(Q_i, Q_j)} \right\} \quad (3)$$

Denklemden n, küme sayısını, S_n kümenin elemanlarının küme merkezine olan uzaklıklarının ortalamasını ve $S_n(Q_i, Q_j)$ iki küme merkezi arasındaki uzaklığı temsil etmektedir. Bu durumda DB değerinin düşük olması kümelerin kendi içinde homojen ve kümelerin birbirlerinden uzak olduğunu belirtmektedir.

Dunn değerinin yüksek, DB değerinin küçük olması küme kalitesinin iyi olduğunu göstermektedir. Değerin büyük ya da küçük olduğunu söyleyebilmek için en az iki senaryo halinde kümeleme yapılması ve her bir senaryo için bu indeks değerlerinin hesaplanması gerekir [34].

Bu çalışmada k-Ortalama ve k-Temsilci yönteminin uygulanmasında kullanılacak k sayısını belirlemek için k=2, k=3, k=4..., k=7 senaryoları için Dunn ve Davies Bouldin indeksleri hesaplanmıştır. 2011, 2012 ve 2013 veri setleri için hesaplanan Dunn ve Davies Bouldin indeksleri Tablo 1 ve Tablo 2’de gösterilmiştir.

Tablo 1. k-Ortalama algoritması için k katsayısının belirlenmesi

	2011 Yılı Veri Seti		2012 Yılı Veri Seti		2013 Yılı Veri Seti	
	Dunn	Davies-Bouldin	Dunn	Davies-Bouldin	Dunn	Davies-Bouldin
k=2	0,027	0,764	0,245	0,543	0,131	0,601
k=3	0,035	0,657	0,087	0,660	0,059	0,632
k=4	0,040	0,642	0,088	0,589	0,089	0,520
k=5	0,034	0,695	0,028	0,659	0,106	0,558
k=6	0,024	0,751	0,036	0,694	0,050	0,635
k=7	0,024	0,721	0,036	0,623	0,043	0,634

3.3.2. AGNES Yöntemi için Parametre Analizi

AGNES metodunda oluşturulacak en uygun küme sayısının belirlenmesinde 2. bölümde anlatılan dendrogram adı verilen ağaç yapıdan yararlanılmaktadır. Dendrogram üzerinde gerçekleşen büyük sıçramalar bize oluşturulacak kümeleri göstermektedir. Ancak kümelemede amaç gerçekten homojen ve birbirlerinden farklı gruplar oluşturmak olduğundan dendrogramın yardımcı bir eksen ile incelenmesi ve homojen grupların belirlenmesi gerekmektedir. Yardımcı eksen ile dendrogram üzerinde ağaç yapının keşiştiği her bir düğüm bir

ÇOK DEĞİŞKENLİ HARİTALAMA İÇİN KÜMELEME YÖNTEMLERİNİN KULLANILMASI

kümeyi ifade etmektedir. Bu çalışmada bu yöntemle 2011, 2012 ve 2013 verileri için küme sayısı 5 olarak belirlenmiştir. AGNES uygulamasından elde edilen kümelerin uygunluğunun belirlenmesinde kullanılan temel yaklaşımlardan biri Cophenetic korelasyon katsayısının (cophenetic correlation coefficient) belirlenmesidir. Cophenetic korelasyon katsayısı, X veri kümesi için hesaplanan benzerlik matrisi $P=\{p_{ij}\}$ ile hiyerarşik kümeleme yöntemine göre elde edilen ağaç diyagramında veri gözlem çiftlerinin ilk defa aynı kümede gruplandığı yakınlık seviyeleri q_{ij} değerlerinden oluşan Cophenetic matrisi $Q=\{q_{ij}\}$ arasındaki benzerliğin bir ölçüsüdür. μ_p ve μ_q sırasıyla $P=\{p_{ij}\}$ ve $Q=\{q_{ij}\}$ 'nin ortalamaları olmak üzere Cophenetic korelasyon katsayısı (CCC) (4) eşitliği ile hesaplanır [2].

Tablo 2. k-Temsilci algoritması için k katsayısının belirlenmesi

	2011 Yılı Veri Seti		2012 Yılı Veri Seti		2013 Yılı Veri Seti	
	Dunn	Davies-Bouldin	Dunn	Davies-Bouldin	Dunn	Davies-Bouldin
k=2	0,005	1,064	0,008	1,250	0,005	1,272
k=3	0,009	0,756	0,017	0,769	0,003	1,117
k=4	0,009	0,746	0,025	0,775	0,002	0,961
k=5	0,023	0,776	0,008	0,807	0,002	1,047
k=6	0,017	0,875	0,008	0,943	0,004	0,904
k=7	0,006	0,911	0,008	0,841	0,004	0,971

$$CCC = \frac{\frac{1}{M} \sum_{i=1}^{n-1} \sum_{j=1}^n p_{ij} q_{ij} - \mu_p \mu_q}{\sqrt{\left(\frac{1}{M} \sum_{i=1}^{n-1} \sum_{j=1}^n p_{ij}^2 - \mu_p^2\right) \left(\frac{1}{M} \sum_{i=1}^{n-1} \sum_{j=1}^n q_{ij}^2 - \mu_q^2\right)}} \tag{4}$$

Burada CCC, [-1,1] arasında değer alır. CCC değerinin 1 olması verideki hiyerarşi yapıdan elde edilen $P=\{p_{ij}\}$ ve $Q=\{q_{ij}\}$ arasında yüksek derecede benzerlik olduğunu gösterir. Bu uygulama kapsamında küme sayısı 5 alınarak 2011,2012 ve 2013 yıllarına ait verilerde bulunan Cophenetic Korelasyon Katsayısı sırasıyla 0,939683, 0,943563 ve 0,937549 olarak hesaplanmıştır. Bu değerler kümeleme sonuçlarımızın kullanılabilir olduğunu göstermektedir, çünkü Romesburg [39] Cophenetic değerinin 0,80 ya da daha büyük olduğunda kullanılan değişkenlerin birbirleriyle ilişkili ve kabul edilebilir olduğunu belirtmiştir.

3.3.3. Değerlendirme

Tablo 1 ve Tablo 2 incelendiğinde k-Ortalama ve k-Temsilci yöntemleri için k küme sayısının 4 veya 5 seçilmesinin uygun olacağı değerlendirilmiştir. AGNES yönteminde ise küme sayısı analizi sonucunda en uygun küme sayısı 5 olarak belirlenmiştir. Oluşturulacak çok değişkenli haritaların karşılaştırılmasının kolay olabilmesi amacıyla çok değişkenli haritaların üretilmesinde 3 yöntem için de küme (sınıf) sayısı 5 olarak seçilmiştir.

3.4. k-Ortalama Yöntemiyle Çok Değişkenli Haritaların Üretilmesi

k-Ortalama yönteminin uygulanmasında RapidMiner yazılımı kullanılmıştır. Bu kapsamda yapılan testler sonucunda en tutarlı sonuçları verecek şekilde k küme sayısı 5, algoritmanın bir kez çalışması esnasında en fazla yapılacak iterasyon sayısı 100, algoritmanın maksimum dönüş sayısı 35 olarak alınmıştır. 2011, 2012 ve 2013 verileri için yöntem ayrı ayrı uygulanmıştır. Kümeleme işlemi sonucu oluşturulan kümelere ait kitle merkezi (centroid) tabloları Tablo 3, Tablo 4 ve Tablo 5’de verilmiştir.

Kümeleme işlemleri sonrasında 4 farklı değeri (motorlu kara taşıtı sayısı, ölümlü ve yaralanmalı trafik kaza sayısı, ölü sayısı ve yaralı sayısı) kullanarak elde edilen sınıflar yardımıyla Türkiye için il bazında trafik kazalarının benzerliklerini gösteren çok değişkenli haritalar üretilmiştir (Şekil 4).

3.5. k-Temsilci Yöntemiyle Çok Değişkenli Haritaların Üretilmesi

k-Temsilci yönteminin uygulanmasında da RapidMiner yazılımından yararlanılmıştır. k-Ortalama algoritmasından farklı olarak k-Ortalama işlem operatörü yerine k-Temsilci işlem operatörü kullanılmıştır. Bu kapsamda k küme sayısı yine 5, algoritmanın maksimum dönüş sayısı da yine 35 olarak alınmıştır. Kümeleme işlemi sonucu oluşturulan kümelere ait kitle merkezi (centroid) tabloları Tablo 6, Tablo 7 ve Tablo 8’de

H. Z. SELVİ, B. ÇAĞLAR

verilmiştir. k-Temsilci yöntemiyle kümeleme işlemleri sonrasında yine aynı 4 farklı değeri kullanarak elde edilen sınıflar yardımıyla yine 2011, 2012 ve 2013 yılları için çok değişkenli haritalar üretilmiştir (Şekil 5).

Tablo 3. 2011 yılı verileri için k-Ortalama yöntemiyle oluşturulan kümelerin kitle merkezi değerleri

K-ORTALAMA KİTLE MERKEZİ TABLOSU (2011)					
	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5
Motorlu Kara Taşıtı Sayısı	2.147.538,5	763.392,7	452.479,6	224.009,6	63.587,3
Ölümlü Yaralanmalı Trafik Kaza Sayısı	12.102,5	6.210,7	3.753	2.057,4	653,1
Ölü Sayısı	195,0	147,7	104,8	68,8	24,2
Yaralı Sayısı	19.319,5	9.957,7	6.330,2	3.865,2	1.337,1

Tablo 4. 2012 yılı verileri için k-Ortalama yöntemiyle oluşturulan kümelerin kitle merkezi değerleri

K-ORTALAMA KİTLE MERKEZİ TABLOSU (2012)					
	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5
Motorlu Kara Taşıtı Sayısı	2.250.907	803.977,3	370.049,0	141.262,4	51.499,9
Ölümlü Yaralanmalı Trafik Kaza Sayısı	13.427	7.607,7	3.544,2	1.500,9	581,9
Ölü Sayısı	226	160,3	75,5	44,0	20,6
Yaralı Sayısı	21.119	11.899	6.015,5	2.857,2	1.155,7

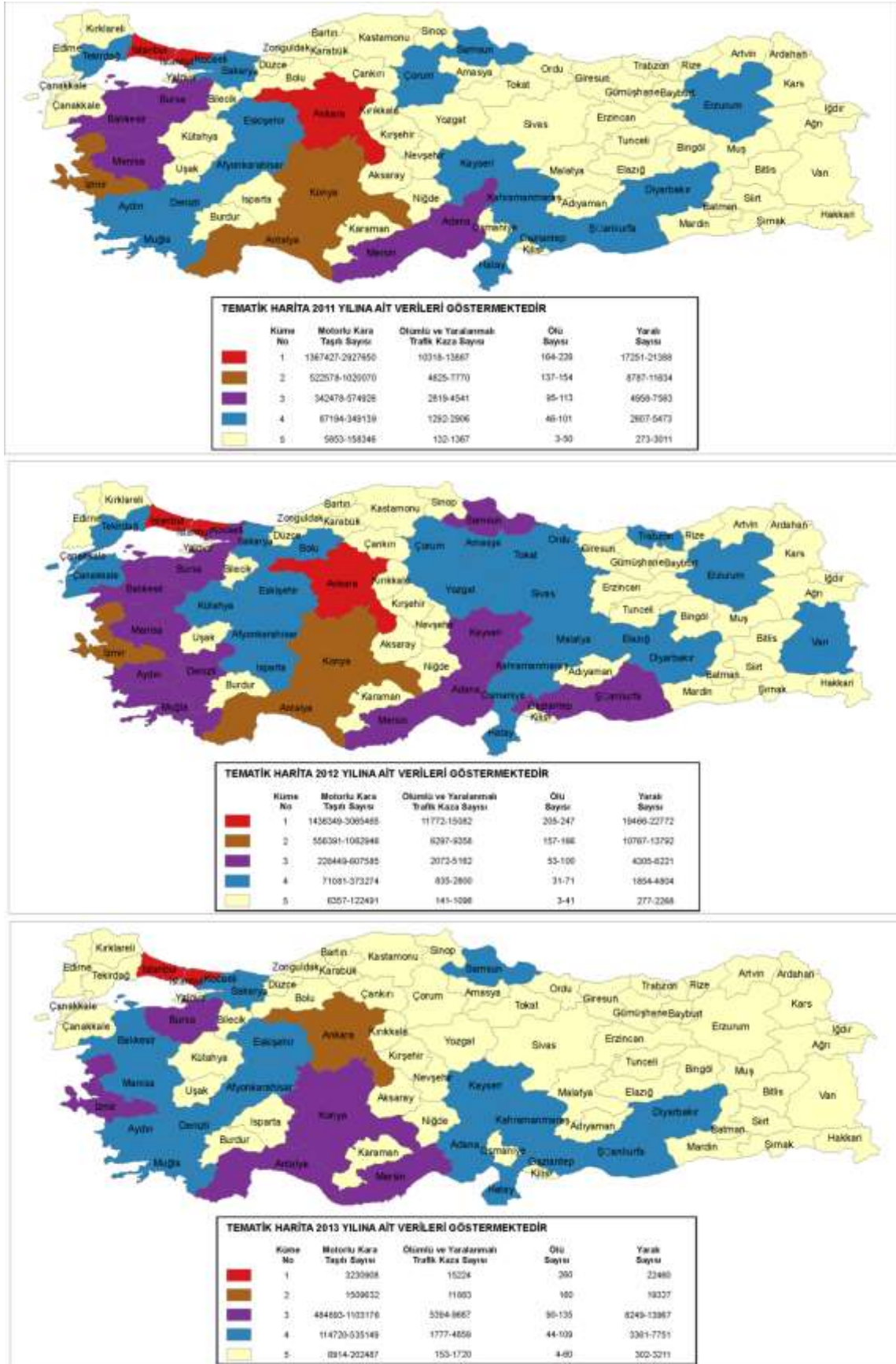
Tablo 5. 2013 yılı verileri için k-Ortalama yöntemiyle oluşturulan kümelerin kitle merkezi değerleri

K-ORTALAMA KİTLE MERKEZİ TABLOSU (2013)					
	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5
Motorlu Kara Taşıtı Sayısı	3.230.908	1.509.632	729.050,0	307.981,9	75.753,8
Ölümlü Yaralanmalı Trafik Kaza Sayısı	15.224	11.883	6.826,6	3.039,6	848,9
Ölü Sayısı	260	160	117,8	70,1	26,1
Yaralı Sayısı	22.460	19.327	10.568,8	5.209,6	1.607,6

Tablo 6. 2011 yılı verileri için k-temsilci yöntemiyle oluşturulan kümelerin kitle merkezi değerleri

K-TEMSİLCİ KİTLE MERKEZİ TABLOSU (2011)					
	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5
Motorlu Kara Taşıtı Sayısı	1.771.715,7	443.392,2	230.032,0	91.045,2	30.875,1
Ölümlü Yaralanmalı Trafik Kaza Sayısı	10.658,3	3.745,8	2.111,1	899,8	362,9
Ölü Sayısı	175,7	112,9	66,7	33,7	13,0
Yaralı Sayısı	16.757,7	6.460,4	3.901,4	1.836,2	754,8

ÇOK DEĞİŞKENLİ HARİTALAMA İÇİN KÜMELEME YÖNTEMLERİNİN KULLANILMASI



Şekil 4. 2011, 2012 ve 2013 yılı için k-ortalama yöntemiyle üretilen çok değişkenli haritalar

H. Z. SELVİ, B. ÇAĞLAR

Tablo 7. 2012 yılı verileri için k-temsili yöntemiyle oluşturulan kümelerin kitle merkezi değerleri

K-TEMSİLCİ KİTLE MERKEZİ TABLOSU (2012)					
	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5
Motorlu Kara Taşıtı Sayısı	1.854.920,0	423.655,3	164.997,7	78.071,2	30.854,4
Ölümlü Yaralanmalı Trafik Kaza Sayısı	12.070,7	4.104,9	1.730,2	831,5	402,4
Ölü Sayısı	203,3	87,4	49,3	31,1	10,8
Yaralı Sayısı	18.676,7	6.842,9	3.288,4	1.657,2	773,2

Tablo 8. 2013 yılı verileri için k-temsili yöntemiyle oluşturulan kümelerin kitle merkezi değerleri

K-TEMSİLCİ KİTLE MERKEZİ TABLOSU (2013)					
	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5
Motorlu Kara Taşıtı Sayısı	2.370.270	498.032,1	228.950,3	141.907,4	64.997,2
Ölümlü Yaralanmalı Trafik Kaza Sayısı	13.553,5	4688	2265,7	1585	742,4
Ölü Sayısı	210	89	84,7	48,6	21,2
Yaralı Sayısı	20.893,5	7515,6	4271,7	2984,5	1398,1

3.6. Birleştirici Hiyerarşik Kümeleme Yöntemiyle (AGNES) Çok Değişkenli Haritaların Üretilmesi

AGNES yönteminin uygulanmasında kümeyi oluşturacak i ve j elemanlarının birbirlerine olan benzerliklerinin belirlenmesinde birbirlerine olan Öklid mesafeleri kullanılmıştır. Kümeyi oluşturacak iki elemanın birbirlerine olan öklid mesafesi;

$$d_{ij} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2} \quad (5)$$

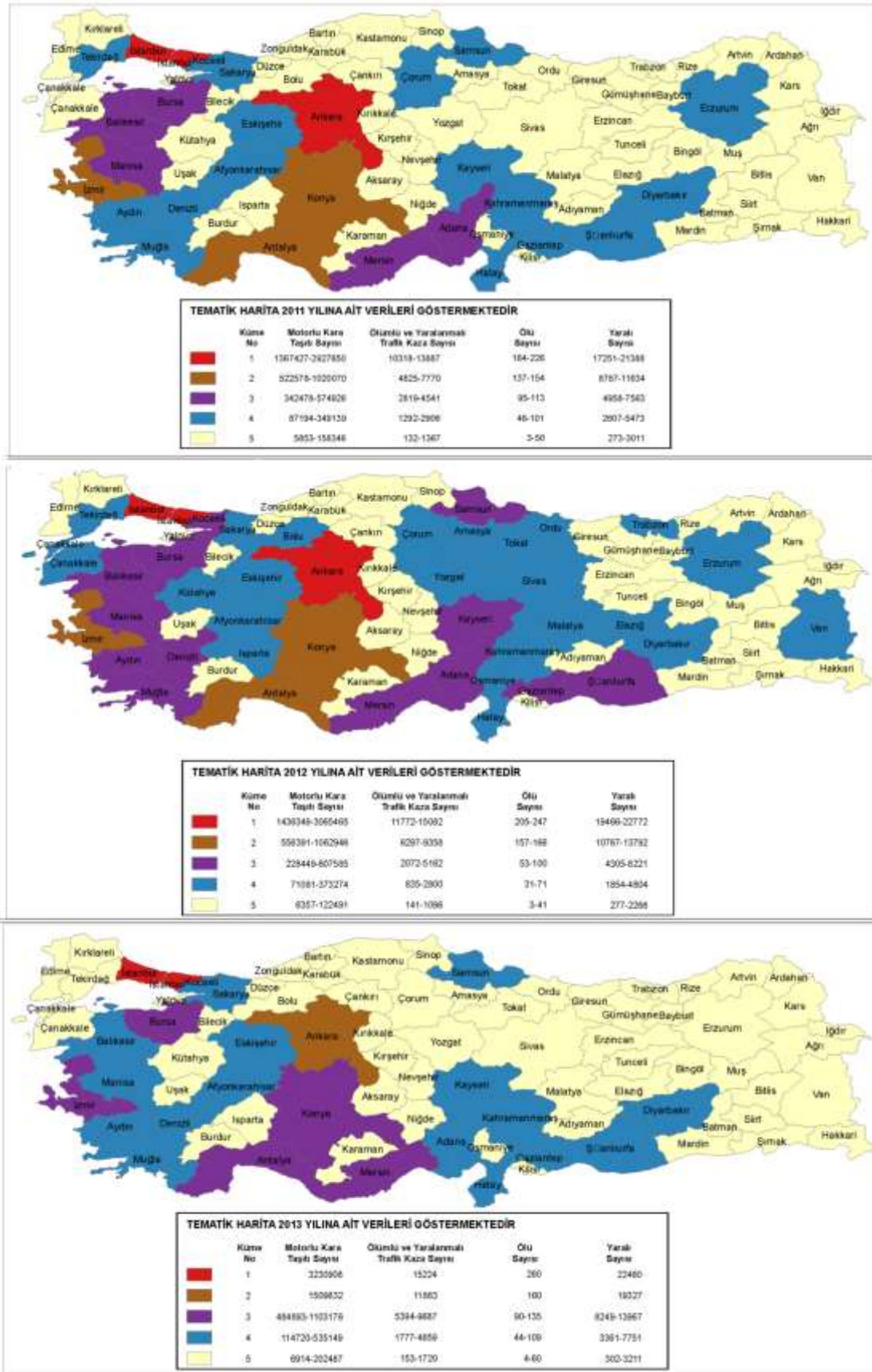
eşitliği ile hesaplanmıştır.

AGNES Hiyerarşik Kümeleme işleminde SPSS yazılımından yararlanılmıştır. Kümeleme işleminin gerçekleşmesinde kullanılan benzeşmezlik matrisi, küme sayısının ve küme elemanlarının belirlenmesinde kullanılacak Dendrogram çizelgesi bu yazılım yardımıyla elde edilmiştir. Yardımcı eksen ile dendrogram üzerinde ağaç yapının keşiştiği her bir düğüm bir kümeyi ifade etmektedir. Bu şekilde 2011, 2012 ve 2013 verileri için 5 küme belirlenmiştir. Kümeleme işlemi sonucu oluşturulan kümelere ait kitle merkezi değerleri Tablo 9, Tablo 10 ve Tablo 11'de verilmiştir. AGNES metoduyla kümeleme işlemleri sonrasında yine aynı 4 farklı değeri kullanarak elde edilen sınıflar yardımıyla yine 2011, 2012 ve 2013 yılları için çok değişkenli haritalar üretilmiştir (Şekil 6).

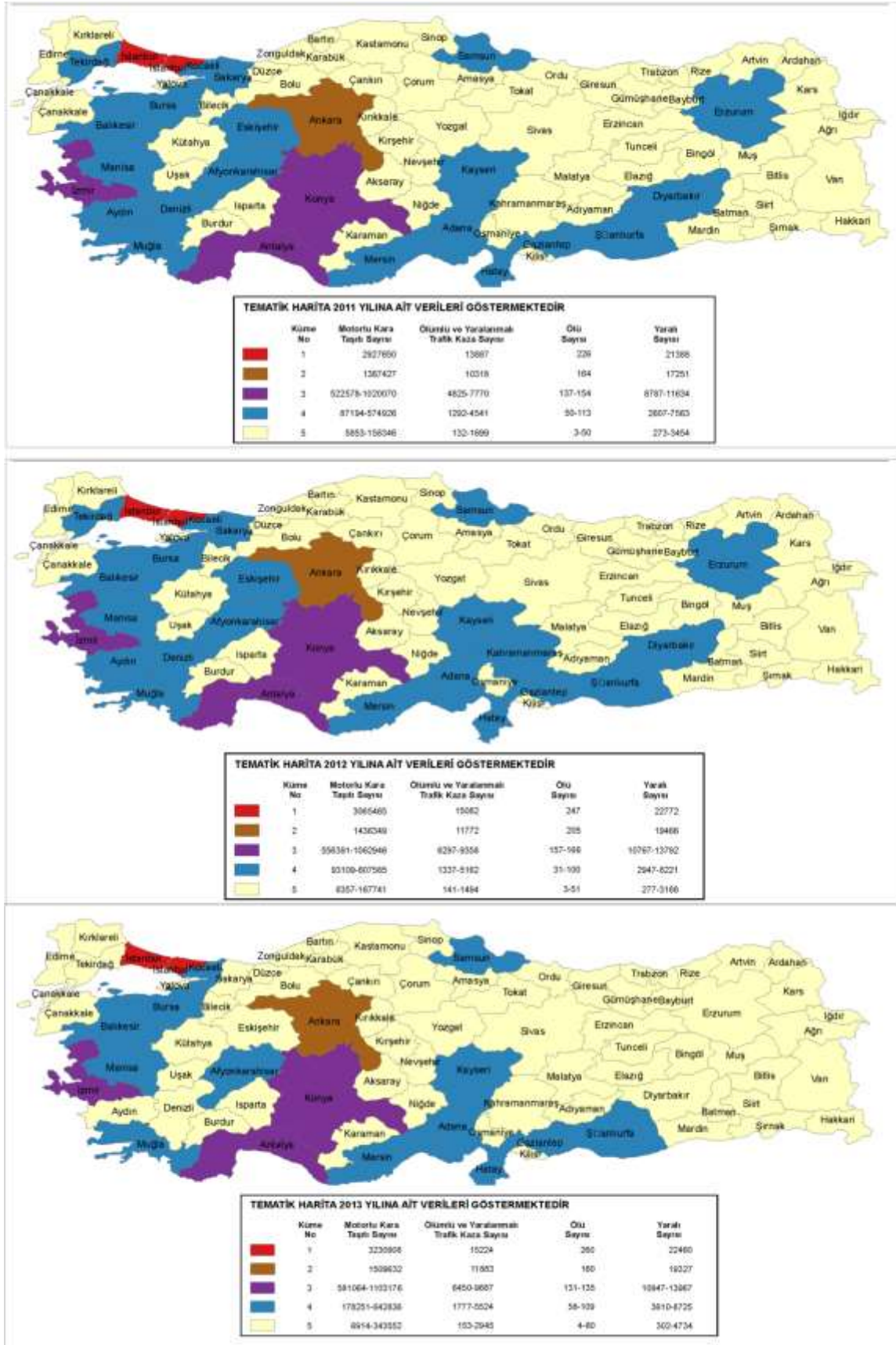
Tablo 9. 2011 yılı verileri için AGNES yöntemiyle oluşturulan kümelerin kitle merkezi değerleri

AGNES KİTLE MERKEZİ TABLOSU (2011)					
	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5
Motorlu Kara Taşıtı Sayısı	2.927.650	1.367.427	763.392,6	289.374,5	66.371,1
Ölümlü Yaralanmalı Trafik Kaza Sayısı	13.887	10.318	6.210,6	2.535,3	683,9
Ölü Sayısı	226	164	147,6	79,9	25,1
Yaralı Sayısı	21.388	17.251	9.957,6	4.553,5	1.401,6

ÇOK DEĞİŞKENLİ HARİTALAMA İÇİN KÜMELEME YÖNTEMLERİNİN KULLANILMASI



Şekil 5. 2011, 2012 ve 2013 yılı için k-temsili yöntemiyle üretilen çok değişkenli haritalar



Şekil 6. 2011, 2012 ve 2013 yılı AGNES metoduyla üretilen çok değişkenli haritalar

ÇOK DEĞİŞKENLİ HARİTALAMA İÇİN KÜMELEME YÖNTEMLERİNİN KULLANILMASI

Tablo 10. 2012 yılı verileri için AGNES yöntemiyle oluşturulan kümelerin kitle merkezi değerleri

AGNES KİTLE MERKEZİ TABLOSU (2012)					
	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5
Motorlu Kara Taşıtı Sayısı	3.065.465	1.436.349	803.977,3	300.970	69.078,1
Ölümlü Yaralanmalı Trafik Kaza Sayısı	15.082	11.772	7.607,6	2.954,7	760,5
Ölü Sayısı	247	205	160,3	66,0	26
Yaralı Sayısı	22.772	19.466	11.899,0	5.130,2	1.498,4

Tablo 11. 2013 yılı verileri için AGNES yöntemiyle oluşturulan kümelerin kitle merkezi değerleri

AGNES KİTLE MERKEZİ TABLOSU (2013)					
	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5
Motorlu Kara Taşıtı Sayısı	3.230.908	1.509.632	839.173,7	382.907,5	90.533,1
Ölümlü Yaralanmalı Trafik Kaza Sayısı	15.224	11.883	7.738,3	3702	997,7
Ölü Sayısı	260	160	133,3	83,1	28,3
Yaralı Sayısı	22.460	19.327	11.956,7	6.182,3	1.854,0

4. SONUÇLAR

Çok değişkenli haritalama kapsamında birden çok özellik ayrı ayrı haritalarda gösterilebileceği gibi, birden çok özellik aynı haritada da gösterilebilir. Her bir özellik için ayrı harita yapılması, haritalardaki objelerin karşılaştırılmasını zorlaştırabilir. Bu nedenle birçok özelliğin aynı haritada gösterildiği yöntemler daha çok tercih edilmektedir. Birçok özelliğin aynı haritada gösterildiği yöntemlerden birisi de farklı özelliklerin etkisini kümeleme analizleriyle belirleyip tematik harita sınıfları oluşturmaktır. Bu kapsamda bu çalışmada Türkiye'deki 2011, 2012 ve 2013'te oluşan trafik kazaları, bu yıllarda trafiğe çıkan araç sayısı, ölümlü ve yaralanmalı trafik kaza sayısı, ölü sayısı ve yaralı sayısı parametreleri dikkate alınarak 3 farklı kümeleme analizi yöntemiyle çok değişkenli haritalar üretilmiştir.

Üretilen haritalar (Şekil 6) ve dendrogram üzerinden belirlenen kümeler incelendiğinde AGNES yöntemiyle üretilen 2011, 2012 ve 2013 haritalarının birbirleriyle oldukça uyumlu olduğu görülmektedir. 2011 verileriyle oluşturulan kümeler ile 2012 verileriyle oluşturulan kümeler karşılaştırıldığında 81 ilden 80 tanesinin, 2011 ve 2013 verileriyle oluşturulan kümeler karşılaştırıldığında ise 81 ilden 74 tanesinin aynı kümeye düştüğü görülmektedir. Bu sonuç AGNES yöntemiyle üretilen çok değişkenli haritaların risk yönetimi açısından da oldukça önemli olduğunu göstermektedir. Çünkü 2011 verileriyle öngörülen risk bölgeleri 2012 ve 2013 verilerinde de doğrulanmıştır.

Hiyerarşik olmayan k-Ortalama ve k-Temsilci kümeleme algoritmaları k giriş parametresine göre n tane nesneyi k tane kümeye bölme mantığıyla çalışmaktadır. Nesnelere birbirlerine benziyor ve diğer kümelerdeki nesnelere benzemiyorlarsa aynı kümeyi oluşturmaktadırlar. Bu algoritmaların uygulanmasındaki en büyük sorun oluşturulacak k küme sayısının belirlenmesidir. Bu da yapılan birkaç uygulama deneyimi veya parametre analizleri ile belirlenebilmektedir. k-Ortalama yöntemiyle 2011 verileriyle oluşturulan kümeler ile 2012 verileriyle oluşturulan kümeler karşılaştırıldığında 81 ilden 59 tanesi, 2011 ve 2013 verileri karşılaştırıldığında ise 81 ilden 71 tanesinin aynı kümeye düştüğü görülmektedir. k-Temsilci yöntemiyle 2011 verileriyle oluşturulan kümeler ile 2012 verileriyle oluşturulan kümeler karşılaştırıldığında 81 ilden 61 tanesi, 2011 ve 2013 verileri karşılaştırıldığında 81 ilden 58 tanesinin aynı kümeye düştüğü görülmektedir. Her iki algoritmanın da kümeleme başarımları benzerlik gösterse de her iki yöntemle oluşturulan kümelerin kitle merkezi çizelgeleri (Tablo 3-8) incelendiğinde k-temsilci algoritmasında kümelerin birbirinden daha iyi ayrıldığı gözlemlenmiştir. Kümeleme çalışmalarındaki amaç küme içi benzerliklerin maksimum, kümeler arası benzerliklerin ise minimum olması olduğundan k-temsilci yönteminin bu veriler için daha iyi sonuç verdiği söylenebilir.

Oluşturulan haritalar incelendiğinde (Şekil 4, 5 ve 6) İstanbul, Ankara, İzmir gibi metropol şehirlerin bu çalışmada ele alınan 4 değişken (motorlu kara taşıtı sayısı, ölümlü ve yaralanmalı trafik kaza sayısı, ölü sayıları ve yaralı sayıları) dikkate alındığında aynı risk seviyesinde ve benzer kümelerde olduğu görülmektedir. Yine Antalya, Konya, Eskişehir, Afyon gibi turistik ve ana ulaşım koridorundaki şehirlerde de oluşan trafik kazalarında benzerlikler görülmektedir. Ayrıca nüfus yoğunluğu düşük olan ve ana yol güzergâhlarında bulunmayan Tunceli, Bitlis, Çankırı vb. şehirlerin de bu 4 parametre dikkate alındığında trafik kazalarının niteliği açısından aynı risk grubunda oldukları gözlenmektedir.

Bu çalışmayla kümeleme yöntemlerinin kullanılmasıyla birden fazla özellik dikkate alınarak farklı mekânsal objelerin benzer yönlerinin ortaya çıkarılabileceği gösterilmiştir. Kümeleme yöntemleriyle oluşturulan çok değişkenli haritalar kullanılarak risk analizi, planlama vb. uygulamalar için önemli katkı sağlayacak mekânsal analizlerin yapılabilceği düşünülmektedir.

KAYNAKLAR

- [1] BUCKLEY, A., Multivariate Mapping. In K. KEMP (Eds.) Encyclopedia of Geographic Information Science (pp. 300-303), 2008.
- [2] SLOCUM, T.A., MCMASTER, R.B., KESSLER, F.C., HOWARD, H.H., Thematic Cartography and Geovisualization, Pearson Education Inc. Third Edition, USA, 2009.
- [3] BREWER, C.A., Color Use Guidelines for Mapping and Visualization, In A.M. MACEACHREN D.R.F TAYLOR (Eds.), Visualization in Modern Cartography (pp. 123-147), 1994.
- [4] METTERNICHT, G., VESTOTT J., "Trivariate Spectral Encoding: A Prototype System for Automated Selection of Colours for Soil Maps Based on Soil Textural Composition", Proceedings of the 21st International Cartographic Conference, Durban, CD, 2003.
- [5] BYRON, J. R., "Spectral Encoding of Soil Texture: A New Visualization Method", GIS/LIS Proceedings, Phoenix, Ariz., 125-132, 1994.
- [6] TERRANTE, V., "Harnessing Natural Textures for Multivariate Visualization", IEEE Computer Graphics and Applications, 20, 6-11, 2000.
- [7] JENKS, G. F., "Pointillism as a Cartographic Technique", The Professional Geographer, 5, 4-6, 1953.
- [8] COX, D.J., "The Art of Scientific Visualization. Academic Computing", 4, 20-22, 32-34, 36-38, 1990.
- [9] ELLSON, R., "Visualization at Work", Academic Computing, 4, 6, 26-28, 54-56, 1990.
- [10] DORLING, D., "The Visualization Of Local Urban Change Across Britain", Environment and Planning B: Planning and Design, 22, 269 -290, 1995.
- [11] GRINSTEIN, G., SIEG, J.C.J., SMITH, S., WILLIAMS M.G., "Visualization for Knowledge Discovery", International Journal of Intelligent Systems, 7, 637-648, 1992.
- [12] HEALEY, C.G., ENNS, J.T., "Large Datasets at A Glance: Combining Textures And Colors In Scientific Visualization", IEEE Transactions on Visualization and Computer Graphics, 5, 2, 145-167, 1999.
- [13] MILLER, J.R., "Attribute Blocks: Visualizing Multiple Continuously Defined Attributes", IEEE Computer Graphics and Applications, 27, 3, 57-69, 2007.
- [14] ZHANG X., PAZNER, M., "The Icon Imagemap Technique for Multivariate Geospatial Data Visualization: Approach and Software System", Cartography and Geographic Information Science, 31, 1, 29-41, 2004.
- [15] NELSON, E. S., GILMARTIN, P. P., An Evaluation Of Multivariate, Quantitative Point Symbols For Maps, In C. H. WOOD, C. P. KELLER (Eds.) Cartographic Design: Theoretical and Practical Perspectives (pp. 191-203), 1996.
- [16] DIBIASE, D., "Designing Animated Maps for A Multimedia Encyclopedia", Cartographic Perspectives, 19, 3-7, 1994.
- [17] NELSON, E.S., "Designing Effective Bivariate Symbols: The Influence of Perceptual Grouping Processes", Cartography and Geographic Information Science, 27, 4, 261-278, 2000.
- [18] GUO, D., GAHEGAN, M., ALAN M. M., BILIAN Z., "Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach", Cartography and Geographic Information Science, 32, 2, 113-132, 2005.
- [19] MENNIS, J., GUO, D., "Spatial Data Mining And Geographic Knowledge Discovery-An Introduction", Computers, Environment and Urban Systems, 33, 403-408, 2009.
- [20] MURRAY, A. T., GRUBESIC, T. H., Exploring Spatial Patterns Of Crime Using Non-Hierarchical Cluster Analysis, In M. Leitner (Eds.) Crime Modeling And Mapping Using Geospatial Technologies, 105-124, Springer, Netherlands, 2013.

ÇOK DEĞİŞKENLİ HARİTALAMA İÇİN KÜMELEME YÖNTEMLERİNİN KULLANILMASI

- [21] GRUBESIC, T. H., WEI, R., MURRAY, A. T., “Spatial Clustering Overview and Comparison: Accuracy, Sensitivity, and Computational Expense”, *Annals of the Association of American Geographers*, 104, 6, 1134-1156, 2014.
- [22] BRIMICOMBE, A.J., “A Dual Approach to Cluster Discovery in Point Event Data Sets”, *Computers, Environment and Urban Systems*, 31, 4-18, 2007.
- [23] ANDERSON T.K, “Kernel Density Estimation and K-Means Clustering to Profile Road Accident Hotspots”, *Accident Analysis and Prevention* 41, 359-364, 2008.
- [24] LU, Q., CHEN, F., HANCOCK, K., “On Path Anomaly Detection in a Large Transportation Network”, *Computers, Environment and Urban Systems*, 33, 6, 448-462, 2009.
- [25] WENG J., WENXIN, Q., XIAOBO, Q., XUEDONG, Y., “Cluster-Based Lognormal Distribution Model For Accident Duration”, *Transportmetrica A: Transport Science*, 11, 4, 345-363,2015.
- [26] GUO, F., FANG, Y., “Individual Driver Risk Analysis Using Naturalistic Driving Data”, 3rd International Conference on Road Safety and Simulation, Indianapolis, USA, 2011.
- [27] YALCIN G., DUZGUN, H.S., “Spatial Analysis of Two-Wheeled Vehicles Traffic Crashes: Osmaniye in Turkey”, *KSCE Journal of Civil Engineering*, 19, 7, 2225-2232, 2015.
- [28] FENG, S, ZHENNING, L., YUSHENG, C., GUOHUI, Z., “Risk Factors Affecting Fatal Bus Accident Severity: Their Impact On Different Types Of Bus Drivers”, *Accident Analysis and Prevention* 86, 29-39, 2016.
- [29] ERDOGAN, S., “Explorative Spatial Analysis Of Traffic Accident Statistics And Road Mortality Among The Provinces Of Turkey”, *Journal of Safety Research* 40, 341-351, 2009.
- [30] MARTINUSSEN, L.M., MØLLER, M., PRATO, C.G., “Assessing The Relationship Between The Driver Behavior Questionnaire And The Driver Skill Inventory: Revealing Sub-Groups Of Drivers”, *Transportation Research Part F* 26, 82-91, 2014.
- [31] DINÇER, E.Ş. Veri Madenciliğinde K-means Algoritması ve Tıp Alanında Uygulanması, Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli, 2006.
- [32] HAN, J., LEE, J.G., KAMBER, M., An Overview of Clustering Methods in Geographic Data Analysis, In Miller H.J., Han H. (Eds.) *Geographic Data Mining and Knowledge Discovery*, Taylor & Francis Group, LLC, 2009.
- [33] HAN, J., KAMBER M., *Data Mining: Concepts and Techniques*. San Francisco, 2006.
- [34] SİLAHTAROĞLU, G., Veri Madenciliği (Kavram ve Algoritmaları). Papatya Yayıncılık, İstanbul, 2013.
- [35] AKIN, Y.K., Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi, Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 2008.
- [36] <http://www.ibm.com/analytics/us/en/technology/spss/> (erişim tarihi 15.12.2016).
- [37] <https://rapidminer.com/> (erişim tarihi 15.12.2016).
- [38] <https://www.arcgis.com/features/index.html> (erişim tarihi 15.12.2016).
- [39] ROMESBURG H.C., *Cluster Analysis for Researchers*, Belmont, CA: Lifetime Learning Publications, 1984.