



TRAFİK KAZA VERİLERİNİN KÜMELEME ANALİZİ YÖNTEMİ İLE İNCELENMESİ

Alper ŞEKERLER

**Temmuz 2008
DENİZLİ**

**TRAFİK KAZA VERİLERİNİN KÜMELEME ANALİZİ
YÖNTEMİ İLE İNCELENMESİ**

**Pamukkale Üniversitesi
Fen Bilimleri Enstitüsü
Yüksek Lisans Tezi
İnşaat Mühendisliği Anabilim Dalı**

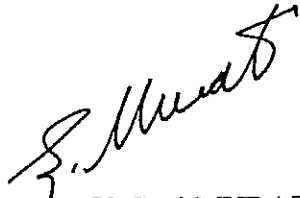
Alper ŞEKERLER

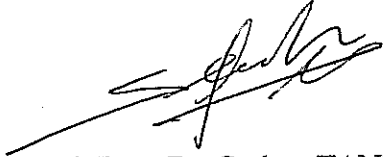
Tez Danışmanı: Doç. Dr. Y. Şazi MURAT


**Temmuz 2008
DENİZLİ**

YÜKSEK LİSANS TEZİ ONAY FORMU

Alper ŞEKERLER tarafından Doç. Dr. Y. Şazi MURAT yönetiminde hazırlanan “Trafik Kaza Verilerinin Kümeleme Analizi Yöntemi İle İncelenmesi” başlıklı tez tarafımızdan okunmuş, kapsamı ve niteliği açısından bir Yüksek Lisans Tezi olarak kabul edilmiştir.


Doç. Dr. Y. Şazi MURAT
Jüri Başkanı


Yrd. Doç. Dr. Serhan TANYEL
Jüri Üyesi


Yrd. Doç. Dr. Mahmut FIRAT
Jüri Üyesi

Pamukkale Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
.../.../..... tarih ve sayılı kararıyla onaylanmıştır.

Prof. Dr. Mehmet Ali SARIGÖL
Müdür

TEŐEKKÜR

Tez alıŐması esnasında her aıdan bana yardımcı olan danıŐmanım Do. Dr. Y. Őazi MURAT' a (PAÜ Mühendislik Fakóltesi İnŐaat MühendisliĐi Bölümü) teŐekkürlerimi arz ederim. DiĐer taraftan Yrd. Do. Dr. Mahmut FIRAT' ın (PAÜ Mühendislik Fakóltesi İnŐaat MühendisliĐi Bölümü) da bana büyük katkısının olduĐunu burada belirtmem gerekir.

105G090 numaralı projeye katkılarında dolay TÜBİTAK'a ve ayrıca Denizli kenti ile ilgili trafik kaza verilerinin derlenmesi aŐamasında yardımları aısından Denizli İl Emniyet MüdürlüĐü' ne teŐekkürlerimi bir bor bilirim.

Beni hayatım boyunca hiç yalnız bırakmayan ve her zaman arkamda olduklarını bildiĐim sevgili annem, babam ve kardeŐime çok teŐekkür ederim.

Alper ŐEKERLER

Bu tezin tasarımı, hazırlanması, yürütülmesi, arařtırmalarının yapılması ve bulgularının analizlerinde bilimsel etięe ve akademik kurallara özenle riayet edildiđini; bu çalıřmanın doğrudan birincil ürünü olmayan bulguların, verilerin ve materyallerin bilimsel etięe uygun olarak kaynak gösterildiđini ve alıntı yapılan çalıřmalara atfedildiđini beyan ederim.

İmza

: 

Öğrenci Adı ve Soyadı: Alper ŞEKERLER

ÖZET

TRAFİK KAZA VERİLERİNİN KÜMELEME ANALİZİ YÖNTEMİ KULLANARAK KÜMELENMESİ

ŞEKERLER, Alper

Yüksek Lisans Tezi, İnşaat Mühendisliği ABD

Tez Yöneticisi: Doç. Dr. Y. Şazi MURAT

Temmuz 2008, 115 Sayfa

Son yıllarda, trafik kazaları tüm dünyada olduğu gibi ülkemizde de çok büyük bir problem haline gelmiştir. Bu problemi çözebilmek için çok sayıda çalışma yapılmış ve yol güvenliğine arttırmaya yönelik projeler gerçekleştirilmiştir. Ancak projelere ayrılan bütçenin sınırlı olması nedeniyle daha düşük maliyetli çözüm yöntemlerinin bulunması gerekli olmuştur. Bu yöntemlerden en önemlisi ise Kara Noktaların belirlenmesi ve iyileştirilmesidir. Geleneksel kara nokta belirleme metodu, her bir kazanın yerinin harita üstünde bir raptiye ile işaretlenmesi ve en çok raptiyenin bulunduğu yerlerin “kara nokta” olarak etiketlenmesidir.

Bu tezde Denizli kentindeki kara noktaları pratik olarak belirlemek için ilk önce trafik kaza verileri kümelere ayrılmıştır. Bunun için klasik ve bulanık kümeleme yöntemleri ele alınmış olup, bu yöntemler ile Denizli kenti için 2004, 2005 ve 2006 yıllarına ait elde edilen trafik kaza verileri kümelere ayrılarak incelenmiştir. Kümeleme sonucunda, küme merkezlerine yakın bölgelerdeki trafik kazalarının daha yoğun olduğu noktalar kara nokta olarak belirlenmiştir.

Klasik ve bulanık kümeleme yöntemleri için trafik kaza verileri, bilgisayar programları yardımı ile iki parametrenin kullanıldığı bir analize tabi tutularak kümelenebilir ve ortaya çıkan sonuçlar yorumlanarak, belirlenen kara noktalar (black spots) için çeşitli çözüm önerileri getirilmiştir. Trafik kaza verileri, koordinat sistemi ve kazaya etki eden faktörler göz önüne alınarak kümelenebilir.

Anahtar Kelimeler: Kümeleme Analizi, K-Ortalamlar Kümeleme Yöntemi, Bulanık C-Ortalamlar Kümeleme Yöntemi, Kara Noktalar, Trafik Kazaları

Doç. Dr. Y. Şazi MURAT

Yrd. Doç. Dr. Serhan TANYEL

Yrd. Doç. Dr. Mahmut FIRAT

ABSTRACT**CLUSTERING OF THE TRAFFIC ACCIDENTS DATA
THROUGH CLUSTER ANALYSIS**

ŞEKERLER, Alper
M. Sc. Thesis in Civil Engineering
Supervisor: Assoc. Prof. Y. Şazi MURAT

July 2008, 115 Pages

In recent years, traffic accidents have become a major problem as have become all over the world, in Turkey as well. To solve this problem, many studies and the projects to develop the road safety are tried. It's necessary to develop the methods with lower cost due to limited budgets. The most important method is to determine the "black spots" and to improve these sections. The conventional black spot identification method is to mark the location of each accident with a pin on a map and to label locations with the most pins as "black spots".

In this thesis, for determining the black spots in Denizli city in a practical way, first of all the traffic accidents data are clustered. Hence, the classic (in this study, k-means method has been used) and the fuzzy clustering methods have been discussed and with these methods, the traffic accidents data of Denizli city for the years of 2004, 2005 and 2006 have been analyzed. At the result of analysis, the spots that have dense of the traffic accidents which are around the cluster centers are determined as "black spot".

For the classic and the fuzzy methods, the traffic accidents data are analyzed by using two parameters through the PC programs and the results of the analysis are commented and are made suggestions for the solution of the black spots. The traffic accidents data are clustered by considering the coordinate system and the factors of the traffic accident.

Keywords: Cluster Analysis, K-Means Clustering Method, Fuzzy C-Means Clustering Method, Black Spots, Traffic Accidents

Assoc. Prof. Y. Şazi MURAT

Asst. Prof. Dr. Serhan TANYEL

Asst. Prof. Dr. Mahmut Fırat

İÇİNDEKİLER

	Sayfa
YÜKSEK LİSANS TEZİ ONAY FORMU.....	i
TEŞEKKÜR.....	ii
BİLİMSEL ETİK SAYFASI.....	iii
ÖZET.....	iv
ABSTRACT.....	v
İÇİNDEKİLER.....	vi
ŞEKİLLER DİZİNİ.....	ix
TABLolar DİZİNİ.....	xi
SİMGE VE KISALTMALAR DİZİNİ.....	xiv
1.GİRİŞ.....	1
1.1 Genel.....	1
1.2 Amaç.....	2
1.3 Kapsam.....	3
2.KÜMELEME ANALİZİ.....	4
2.1 Genel.....	4
2.2 Kümeleme Metotları.....	7
2.2.1 Bölümlemeli kümeleme metodu.....	8
2.2.1.1 Klasik bölümlendirme metodu.....	8
2.2.1.1.1 K-ortalamalar metodu.....	8
2.2.1.1.2 K-medoidler metodu.....	9
2.2.1.2 Modern bölümlendirme metodu.....	11
2.2.1.2.1 Hard k-ortalamalar metodu	11
2.2.1.2.2 Bulanık c-ortalamalar metodu.....	12
2.2.1.2.2.1 Fanny algoritması.....	15
2.2.1.2.2.2 Doğrulama.....	17
2.2.2 Hiyerarşik (aşamalı) kümeleme metodu.....	18
2.2.2.1 Toplayıcı (agglomerative) hiyerarşik kümeleme metodu.....	20
2.2.2.1.1 Tek bağlantılı kümeleme yöntemi (TekBKY).....	21
2.2.2.1.2 Tam bağlantılı kümeleme yöntemi (TBKY).....	22
2.2.2.1.3 Ortalama bağlantılı kümeleme yöntemi (OrtBKY).....	24
2.2.2.1.4 Ortalama grup bağı.....	24
2.2.2.1.5 Ward'ın hiyerarşik kümeleme metodu.....	25
2.2.2.2 Bölen metot.....	26
2.2.2.2.1 DIANA.....	26
2.2.2.2.1.1 Bölen katsayısı (BK).....	27
2.2.3 Yoğunluk tabanlı kümeleme metodu.....	28
2.2.3.1 DBSCAN.....	28
2.2.3.1.1 Bir noktanın eps_komşuluğu.....	28
2.2.3.1.2 Direkt yoğunluk ulaşılabilirlik.....	29
2.2.3.1.3 Yoğunluk ulaşılabilirlik.....	29
2.2.3.1.4 Yoğunluk bağıllık.....	29
2.2.3.2 OPTICS.....	31
2.2.3.2.1 Yoğunluk tabanlı küme sıralaması.....	31
2.2.3.2.2 Bir nesnenin öz uzaklığı.....	32
2.2.3.2.3 Ulaşılabilirlik mesafesi.....	32
2.2.3.2.3 FJP metodu (bulanık düğüm noktaları).....	35

2.2.4 Grid (ızgara) bazlı kümeleme metodu.....	35
2.2.4.1 STING (istatistiksel bilgi gridi).....	35
2.2.4.2 WaveCluster (wavelet dönüşümü kullanarak kümeleme).....	36
2.2.4.3 CLIQUE (yüksek boyutlu alanda kümeleme).....	37
3. ÖNCEKİ ÇALIŞMALAR.....	39
3.1 Genel.....	39
3.2 Daha Önce Yapılmış Çalışmalar.....	40
4. ÇALIŞMA ALANI VE VERİ.....	45
4.1 Genel.....	45
4.2 Çalışma Alanı ve Kullanılabilir Veri.....	45
4.3 NCSS Hakkında.....	47
4.3.1 NCSS veri penceresi.....	48
4.3.2 NCSS işlem penceresi.....	48
4.3.3 NCSS çıktı penceresi.....	48
4.3.4 NCSS ile k-ortalamlar kümeleme analizi.....	49
5. TRAFİK KAZA VERİLERİNİN ANALİZİ.....	51
5.1 Genel İstatistikler.....	51
5.2 Konum (x, y) Verisinin K-Ortalamlar ile Kümeleme Analizi.....	58
5.3 Konum (x, y) Verisinin Bulanık Kümeleme Analizi.....	62
5.4 K-Ortalamlar Metodu ile Diğer Kümeleme Analizleri.....	67
5.4.1 Yol türü – yol bölünme.....	68
5.4.2 Yol türü – oluşumuna göre kaza türü.....	70
5.4.3 Yol türü – hava durumu.....	71
5.4.4 Yol türü – gün durumu.....	72
5.4.5 Yol türü – yolda yön.....	73
5.4.6 Yol türü – yolun kaplama cinsi.....	74
5.4.7 Yol bölünme – oluşumuna göre kaza türü.....	75
5.4.8 Yol bölünme – hava durumu.....	76
5.4.9 Yol bölünme – gün durumu.....	77
5.4.10 Yol bölünme – yolun kaplama cinsi.....	78
5.4.11 Yerleşim durumu – oluşumuna göre kaza türü.....	79
5.4.12 Yerleşim durumu – hava durumu.....	80
5.4.13 Yerleşim durumu – gün durumu.....	81
5.4.14 Yerleşim durumu – yolda yön.....	82
5.4.15 Yerleşim durumu – yolun kaplama cinsi.....	83
5.4.16 Hava durumu – oluşumuna göre kaza türü.....	84
5.4.17 Hava durumu – yolun kaplama cinsi.....	85
5.5 Alt Kümeleme Analizleri.....	86
5.5.1 Karşılıklı çarpışma.....	86
5.5.2 Arkadan çarpma.....	88
5.5.3 Yandan çarpma veya yandan çarpışma.....	90
5.5.4 Duran araca çarpma.....	92
5.5.5 Sabit cisme çarpma.....	94
5.5.6 Yayaya çarpma.....	96
5.5.7 Hayvana çarpma.....	98
5.5.8 Devrilme.....	99
5.5.9 Yoldan çıkma.....	101
5.5.10 Araçtan düşen insan.....	103
5.5.11 Araçtan düşen cisim.....	104
6. SONUÇLAR VE ÖNERİLER.....	106

KAYNAKLAR.....	112
ÖZGEÇMİŞ.....	115

ŞEKİLLER DİZİNİ

	Sayfa
Şekil 1.1 Yolcu taşımacılığında ulaşım türlerini kullanma oranları (%) (Ege 1999).....	2
Şekil 2.1 Kümeleme sürecinin adımları (Güler 2006).....	6
Şekil 2.2 k-medoidler metoduyla kümeleme örneği (Avcı 2005).....	10
Şekil 2.3 Hiyerarşik kümeleme için dendogram.....	19
Şekil 2.4 Ham veri (WEB_1: http://en.wikipedia.org/wiki/Data_clustering).....	20
Şekil 2.5 Geleneksel temsili.....	21
Şekil 2.6 Tek bağlantılı kümeleme modeli.....	22
Şekil 2.7 Tam bağlantılı kümeleme modeli.....	23
Şekil 2.8 Ortalama bağlantılı kümeleme modeli.....	24
Şekil 2.9 (a) Yoğunluk ulaşılabilirlik ve (b) yoğunluk bağlılık.....	30
Şekil 2.10 STING kümeleme için hiyerarşik yapı.....	36
Şekil 3.1 Kart-sınıflama görevi için örnek sayfa kartı.....	42
Şekil 4.1 Denizli kenti 2004, 2005 ve 2006 yıllarına ait trafik kaza verisinin harita üzerindeki dağılımı.....	47
Şekil 4.2 NCSS veri penceresi.....	48
Şekil 4.3 NCSS işlem penceresi.....	48
Şekil 4.4 NCSS çıktı penceresi.....	49
Şekil 4.5 NCSS minimum iterasyon çıktısı.....	49
Şekil 4.6 NCSS k-ortalamlar kümeleme analizi işlem penceresi.....	50
Şekil 5.1 Yıllara göre kaza dağılımı.....	51
Şekil 5.2 Aylara göre kaza dağılımı.....	52
Şekil 5.3 Haftanın günlere göre kaza dağılımı.....	53
Şekil 5.4 Saatlere göre kaza dağılımı.....	54
Şekil 5.5 Yol türüne göre kaza dağılımı.....	55
Şekil 5.6 Yol bölünme durumuna göre kaza dağılımı.....	55
Şekil 5.7 Oluşumuna göre kaza dağılımı.....	56
Şekil 5.8 Hava durumuna göre kaza dağılımı.....	57
Şekil 5.9 Gün durumuna göre kaza dağılımı.....	57
Şekil 5.10 Yolun kaplama cinsine göre kaza dağılımı.....	58
Şekil 5.11 GIS' de Denizli haritasındaki örnek yerleştirilmiş trafik kaza verisi.....	59
Şekil 5.12 Konum verisinin k-ortalamlar küme analizi sonucu küme grafiği.....	61
Şekil 5.13 Dunn indeksi (DI) ve alternatif Dunn indeksi (ADI).....	63
Şekil 5.14 Ayırma indeksi (PC), ayrıştırma indeksi (S), Xie ve Beni indeksi (XB) ..	63
Şekil 5.15 Amaç fonksiyonu grafiği.....	64
Şekil 5.16 Konum verisinin bulanık kümeleme analizi sonucu küme grafiği.....	66
Şekil 5.17 Yol türü - yol bölünme küme grafiği.....	69
Şekil 5.18 Yol türü - oluşumuna göre kaza türü, küme grafiği.....	70
Şekil 5.19 Yol türü - hava durumu, küme grafiği.....	71
Şekil 5.20 Yol türü - gün durumu, küme grafiği.....	72
Şekil 5.21 Yol türü - yolda yön, küme grafiği.....	73
Şekil 5.22 Yol türü - yolun kaplama cinsi, küme grafiği.....	74
Şekil 5.23 Yol bölünme - oluşumuna göre kaza türü, küme grafiği.....	75
Şekil 5.24 Yol bölünme - hava durumu, küme grafiği.....	76
Şekil 5.25 Yol bölünme - gün durumu, küme grafiği.....	77
Şekil 5.26 Yol bölünme - yolun kaplama cinsi, küme grafiği.....	78

Şekil 5.27 Yerleşim durumu - oluşumuna göre kaza türü, küme grafiği.....	79
Şekil 5.28 Yerleşim durumu - hava durumu, küme grafiği.....	80
Şekil 5.29 Yerleşim durumu - gün durumu, küme grafiği.....	81
Şekil 5.30 Yerleşim durumu - yolda yön, küme grafiği.....	82
Şekil 5.31 Yerleşim durumu - yolun kaplama cinsi, küme grafiği.....	83
Şekil 5.32 Hava durumu - oluşumuna göre kaza türü, küme grafiği.....	84
Şekil 5.33 Hava durumu - yolun kaplama cinsi, küme grafiği.....	85
Şekil 5.34 Koordinat verisi (karşılıklı çarpışma), küme grafiği.....	87
Şekil 5.35 Koordinat verisi (arkadan çarpma), küme grafiği.....	89
Şekil 5.36 Koordinat verisi (yandan çarpma), küme grafiği.....	91
Şekil 5.37 Koordinat verisi (duran araca çarpma), küme grafiği.....	93
Şekil 5.38 Koordinat verisi (sabit cisme çarpma), küme grafiği.....	95
Şekil 5.39 Koordinat verisi (yayaya çarpma), küme grafiği.....	97
Şekil 5.40 Koordinat verisi (hayvana çarpma), küme grafiği.....	98
Şekil 5.41 Koordinat verisi (devrilme), küme grafiği.....	100
Şekil 5.42 Koordinat verisi (yoldan çıkma), küme grafiği.....	102
Şekil 5.43 Koordinat verisi (araçtan düşen insan), küme grafiği.....	103
Şekil 5.44 Koordinat verisi (araçtan düşen cisim), küme grafiği.....	104

TABLOLAR DİZİNİ

	Sayfa
Tablo 3.1 Çalışma sonucu elde edilen çarpışma ve yayaya çarpma bakımından illerin kümelenmesi (Karpat ve Yılmaz 1997).....	40
Tablo 3.2 Çalışmada elde edilen kümeleme analizi sonuçları (Cebeci 2004).....	43
Tablo 3.3 Çalışma sonucu elde edilen kümeleme tablosu (Turanlı vd. 2006).....	44
Tablo 5.1 Yıllara göre kaza dağılımı.....	51
Tablo 5.2 Aylara göre kaza dağılımı.....	52
Tablo 5.3 Haftanın günlerine göre kaza dağılımı.....	52
Tablo 5.4 Saatlere göre kaza dağılımı.....	53
Tablo 5.5 Yol türüne göre kaza dağılımı.....	54
Tablo 5.6 Yol bölünme durumuna göre kaza dağılımı.....	55
Tablo 5.7 Oluşumuna göre kaza dağılımı.....	56
Tablo 5.8 Hava durumuna göre kaza dağılımı.....	56
Tablo 5.9 Gün durumuna göre kaza dağılımı.....	57
Tablo 5.10 Yolun kaplama cinsine göre kaza dağılımı.....	58
Tablo 5.11 Konum verisinin k-ortalamlar kümeleme analizi sonucu minimum iterasyonu.....	59
Tablo 5.12 Konum verisinin k-ortalamlar kümeleme analizi sonucu iterasyonu.....	60
Tablo 5.13 Konum verisinin k-ortalamlar kümeleme analizi sonucu küme merkezleri.....	60
Tablo 5.14 Konum verisinin k-ortalamlar kümeleme analizi sonucu uzaklıklar matrisi.....	61
Tablo 5.15 Konum verisinin k-ortalamlar kümeleme analizi sonucu belirlenen kara noktalar.....	61
Tablo 5.16 Doğrulama katsayıları.....	62
Tablo 5.17 Amaç fonksiyonu iterasyon adımları.....	64
Tablo 5.18 Eleman üyelik dereceleri.....	65
Tablo 5.19 Konum verisinin bulanık kümeleme analizi sonucu küme merkezleri.....	65
Tablo 5.20 Konum verisinin bulanık kümeleme sonucu belirlenen kara noktalar.....	66
Tablo 5.21 Kümeleme parametreleri.....	68
Tablo 5.22 Yol türü - yol bölünme, iterasyon.....	69
Tablo 5.23 Yol türü - yol bölünme küme merkezleri.....	69
Tablo 5.24 Yol türü - oluşumuna göre kaza türü, iterasyon.....	70
Tablo 5.25 Yol türü - oluşumuna göre kaza türü, küme merkezleri.....	70
Tablo 5.26 Yol türü - hava durumu, iterasyon.....	71
Tablo 5.27 Yol türü - hava durumu, küme merkezleri.....	71
Tablo 5.28 Yol türü - gün durumu, iterasyon.....	72
Tablo 5.29 Yol türü - gün durumu, küme merkezleri.....	72
Tablo 5.30 Yol türü - yolda yön, iterasyon.....	73
Tablo 5.31 Yol türü - yolda yön, küme merkezleri.....	73
Tablo 5.32 Yol türü - yolun kaplama cinsi, iterasyon.....	74
Tablo 5.33 Yol türü - yolun kaplama cinsi, küme merkezleri.....	74
Tablo 5.34 Yol bölünme - oluşumuna göre kaza türü, iterasyon.....	75
Tablo 5.35 Yol bölünme - oluşumuna göre kaza türü, küme merkezleri.....	75
Tablo 5.36 Yol bölünme - hava durumu, iterasyon.....	76
Tablo 5.37 Yol bölünme - hava durumu, küme merkezleri.....	76
Tablo 5.38 Yol bölünme - gün durumu, iterasyon.....	77

Tablo 5.39 Yol bölünme - gün durumu, küme merkezleri.....	77
Tablo 5.40 Yol bölünme - yolun kaplama cinsi, iterasyon.....	78
Tablo 5.41 Yol bölünme - yolun kaplama cinsi, küme merkezleri.....	78
Tablo 5.42 Yerleşim durumu - oluşumuna göre kaza türü, iterasyon.....	79
Tablo 5.43 Yerleşim durumu - oluşumuna göre kaza türü, küme merkezleri.....	79
Tablo 5.44 Yerleşim durumu - hava durumu, iterasyon.....	80
Tablo 5.45 Yerleşim durumu - hava durumu, küme merkezleri.....	80
Tablo 5.46 Yerleşim durumu - gün durumu, iterasyon.....	81
Tablo 5.47 Yerleşim durumu - gün durumu, küme merkezleri.....	81
Tablo 5.48 Yerleşim durumu - yolda yön, iterasyon.....	82
Tablo 5.49 Yerleşim durumu - yolda yön, küme merkezleri.....	82
Tablo 5.50 Yerleşim durumu - yolun kaplama cinsi, iterasyon.....	83
Tablo 5.51 Yerleşim durumu - yolun kaplama cinsi, küme merkezleri.....	83
Tablo 5.52 Hava durumu - oluşumuna göre kaza türü, iterasyon.....	84
Tablo 5.53 Hava durumu - oluşumuna göre kaza türü, küme merkezleri.....	84
Tablo 5.54 Hava durumu - yolun kaplama cinsi, iterasyon.....	85
Tablo 5.55 Hava durumu - yolun kaplama cinsi, küme merkezleri.....	85
Tablo 5.56 Koordinat verisi (karşılıklı çarpışma), iterasyon.....	86
Tablo 5.57 Koordinat verisi (karşılıklı çarpışma), küme merkezleri.....	87
Tablo 5.58 Karşılıklı çarpışmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar.....	88
Tablo 5.59 Koordinat verisi (arkadan çarpma), iterasyon.....	88
Tablo 5.60 Koordinat verisi (arkadan çarpma), küme merkezleri.....	88
Tablo 5.61 Arkadan çarpmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar.....	89
Tablo 5.62 Koordinat verisi (yandan çarpma), iterasyon.....	90
Tablo 5.63 Koordinat verisi (yandan çarpma), küme merkezleri.....	90
Tablo 5.64 Yandan çarpmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar.....	91
Tablo 5.65 Koordinat verisi (duran araca çarpma), iterasyon.....	92
Tablo 5.66 Koordinat verisi (duran araca çarpma), küme merkezleri.....	92
Tablo 5.67 Duran araca çarpmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar.....	93
Tablo 5.68 Koordinat verisi (sabit cisme çarpma), iterasyon.....	94
Tablo 5.69 Koordinat verisi (sabit cisme çarpma), küme merkezleri.....	94
Tablo 5.70 Sabit cisme çarpmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar.....	95
Tablo 5.71 Koordinat verisi (yayaya çarpma), iterasyon.....	96
Tablo 5.72 Koordinat verisi (yayaya çarpma), küme merkezleri.....	96
Tablo 5.73 Yayaya çarpmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar.....	97
Tablo 5.74 Koordinat verisi (hayvana çarpma), iterasyon.....	98
Tablo 5.75 Koordinat verisi (hayvana çarpma), küme merkezleri.....	98
Tablo 5.76 Hayvana çarpmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar.....	99
Tablo 5.77 Koordinat verisi (devrilme), iterasyon.....	99
Tablo 5.78 Koordinat verisi (devrilme), küme merkezleri.....	100
Tablo 5.79 Devrilmeye ait konum verisinin kümeleme analizi sonucu.....	100
Tablo 5.80 Koordinat verisi (yoldan çıkma), iterasyon.....	101
Tablo 5.81 Koordinat verisi (yoldan çıkma), küme merkezleri.....	101

Tablo 5.82 Yoldan çıkmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar.....	102
Tablo 5.83 Koordinat verisi (araçtan düşen insan), iterasyon.....	103
Tablo 5.84 Koordinat verisi (araçtan düşen insan), küme merkezleri.....	103
Tablo 5.85 Araçtan düşen insana ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar.....	104
Tablo 5.86 Koordinat verisi (araçtan düşen cisim), iterasyon.....	104
Tablo 5.87 Koordinat verisi (araçtan düşen cisim), küme merkezleri.....	104
Tablo 5.88 Araçtan düşen cisme ait konum verisinin kümeleme analizi sonucu.....	105
Tablo 6.1 Kaza türlerine göre tespit edilen kara noktalar.....	110
Tablo 6.2 Geometrisi değişen kavşaklardaki yıllık kaza sayıları.....	111

SİMGE VE KISALTMALAR DİZİNİ

BK	Bölen katsayısı
CE	Classification Entropy (Sınıflandırma Entropisi)
CBS	Coğrafi Bilgi Sistemi
C_i	i. küme merkezi
CLIQUE	Clustering High Dimensional Space (Yüksek Boyutlu Alanda Kümeleme)
d	Merkez ile veri arasındaki mesafe
D	Uzaklık matrisi
DBSCAN	Density Based Spatial Clustering of Application with Noise
DI	Dunn's Index (Dunn İndeksi)
d_{ij}	i. merkez c_i ile j. veri noktası arasındaki öklit mesafesi
Eps	Uzaklık değeri
ESS	Error Sum-of-Squares Criterion (Hata Kareler Toplamı Kriteri)
F_k	Dunn ayrıştırma katsayısı
FJP	Fuzzy Joint Points
GIS	Geographic Information System
J	n verinin kendi küme merkezlerinden olan toplam uzaklığı
m	Ağırlıklandırma faktörü
MinPts	Bir doğal sayı
N_r	r kümesinin boyutu
N_s	s kümesinin boyutu
$N_{\epsilon}(p)$	p'nin Eps_Komşuluğu
O_i	i kümesinin medoidi
O_j	j kümesinin medoidi
OPTICS	Ordering Points to Identify the Clustering Structure
$O_{rasgele}$	Seçilen ve medoid adayı olan bir nesne
OrtBKY	Ortalama Bağlantılı Kümeleme Yöntemi
p	Bir veri tabanında nesne
PC	Partition coefficient (Ayrırma Katsayısı)
R	Korelasyon katsayısı
S	Separation Index (Ayrıştırma indeksi)
SC	Partition Index (Ayrırma İndeksi)
Sim	Benzerlik matrisi
Simij	Benzerlik matrisi elemanları
s(i)	Siluet Katsayısı (silhouette coefficient)
STING	Statistical Information Grid (İstatistiksel Bilgi Gridi)
TekBKY	Tek Bağlantılı Kümeleme Yöntemi
T_{rs}	r ve s kümesi arasındaki tüm çiftlerin mesafeleri toplamı
U	Üyelik matrisi
u_{ij}	j. veri noktasının i. gruba ait üyelik matrisi elemanı
u_{iv}	i. birimin v kümesine ne kadar ait olduğunu gösteren üyelik fonksiyonu
X	Veri matrisi
XB	Xie and Beni's Index (Xie ve Beni İndeksi)
x_j	1, ..., n veri seti için k-ortalamlar algoritması

1.GİRİŞ

1.1 Genel

Gelişmekte olan dünyada özellikle son elli yılda nüfusun hızla artması nedeniyle trafik de kötü yönde bir değişim geçirmiştir. Dünya giderek küreselleşmiş ve bu küreselleşmenin en önemli etkeni de trafik olmuştur. Trafikteki bu büyümenin bazı kötü etkileri, örneğin, trafik tıkanıklığı ve hava kirliliği, doğrudan gözle görülür bir biçimde anlaşılabilir. Trafik tıkanıklığı ve hava kirliliği, doğrudan gözle görülür bir biçimde anlaşılabilir.

Trafiğin artmasına bağlı olarak trafik sorunları da artmaya başlamıştır. Kazalar, otoyol yapımı, araç üretimi, güvenliğin teknolojik açıdan sağlanması ve kazaların önlenmesi bu alanda başlıca sorunlardan bazılarıdır. Trafik kazalarının ana nedenlerinin başında söz konusu hizmetlerin eksikliği gelmektedir. Ülkemizde bu eksiklikler arasında; bölünmüş yollarımızın azlığı, yol yatay ve düşey işaretlerinin eksikliği, her gün artan otopark ihtiyacı ve bu ihtiyacın giderilememesi en çok göze çarpan unsurlardır. Bunların dışında, çocuklarımız için oyun sahalarının yetersizliği, taşıtların bakım ve teknik muayenelerinin eksikliği de olumsuzluklar arasında sayılabilir. Ayrıca gelişmiş ülkelerde trafiğin değişmez unsuru sayılan trafik kontrol merkezleri açısından ülkemiz oldukça fakirdir. Büyük şehirlerimizin trafik düzenlenmesinde, kapalı devre kameraları ile trafiği düzenleyen trafik kontrol merkezleri ya yetersiz ya da yoktur.

Trafik kazaları rasgele oluşmalarına rağmen, sıklığı birçok faktör tarafından etkilenmektedir. Kazalar istatistiki bir değer olarak ele alındığında, hiç kimse oluşabilecek bir trafik kazasının yerini ve zamanını tahmin edemez. Kazaya etki eden faktörler başlıca insan, taşıt, yol ve çevresi başlıkları altında toplanabilir. Trafik kazalarından az zarar görülmesi, alınan tedbirlerin etkinliğine bağlıdır. Başlıca tedbir alanları;

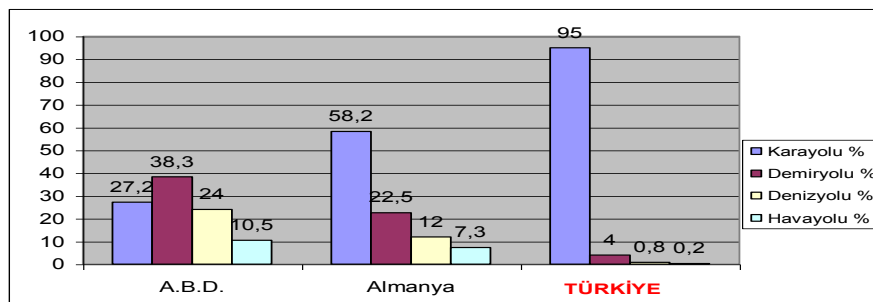
- *Trafik Mühendisliği Faaliyetleri:* Bölünmüş yol yapımı, otopark, oyun alanı, yaya kaldırımı, üst ve alt geçit inşaatı, güvenli taşıt üretilmesi, trafik kontrol merkezleri ile trafik düzenlenmesi vb.
- *Eğitim Faaliyetleri:* Trafik güvenliğinin sağlanması amacıyla başta yolu kullanan sürücü, yolcu ve yayalar olmak üzere, trafikle yetkili yetkisiz yani halkın eğitilmesi, bilgilendirilmesi, motive edilmesi vb.

- *Denetim ve Yasal Düzenleme Faaliyetleri:* Kural ihlallerini önlemeye yönelik dürüst, sık ve teknolojidenden yararlanılarak trafik kontrolleri ile trafik güvenliğini geliştirici ve uygulanabilir yasal düzenlemelerin yapılması vb.
- *İlk ve Acil Sağlık Hizmetleri:* Kaza sonrası mağdurlardan hafif yaralananın ağır hale gelmemesi, ağır yaralının ölmesini engellemek için süratli ve kaliteli hizmet veren uzman personel ve donanımlı ambulansın kaza yerinde görev yapması vb. dir.

1.2 Amaç

Karayolu trafik kazaları çok ciddi bir sorun olup, önlenmesi için birçok alanda etkili faaliyet yapılması ve ilgili çalışmaların verimli koordine edilmesi gereklidir. Türkiye’de meydana gelen trafik kazalarının birden çok faktörü olduğu gibi trafik mühendisliği, eğitimi, düzenlemesi ve denetlemesi konusundaki yetersizlikler, kaza sonrası ölümlerin sayısındaki fazlalık sebebi ile de, ilk ve acil sağlık hizmetlerindeki eksiklikler de bunlara ek olarak göz önünde bulundurulmalıdır.

Türkiye’deki trafik kazalarının ana nedenlerinin başında yük ve yolcu taşımacılığında karayoluna ağırlık verilmesi gelmektedir. Yapılan bilimsel araştırma neticelerine göre; karayolu demiryoluna göre 18, havayoluna göre de 27 kat daha tehlikelidir. Ülkemizde ve bazı yabancı ülkelerde yolcu taşımacılığında ulaşım türlerini kullanma oranları Şekil 1.1’ de verilmiştir. Karayolu taşımacılığımız çok yüksek olmasına rağmen otoyol ağımızın uzunluğu çok azdır. Otoyollar, devlet yollarına göre 4 kat daha güvenlidir.



Şekil 1.1 Yolcu taşımacılığında ulaşım türlerini kullanma oranları (%)

Bu çalışmanın genel amacı kara yollarında *trafik kazalarının yoğun görüldüğü noktaların (kara noktaların) yerlerinin ve sayılarının kümeleme analizi ile tespit edilerek, bu noktaların iyileştirilmesi ve azaltılması yönünde çeşitli çözüm önerilerinin*

getirilmesidir. Özel amacı ise hangi tür kazaların daha çok hangi noktalarda oluştuğunun tespit edilmesidir.

1.3 Kapsam

Bu çalışma altı bölümden oluşmaktadır.

İkinci bölümde, kümeleme analizi ve kümeleme metotları ile ilgili ayrıntılı bilgiler verilmiştir.

Üçüncü bölümde, kümeleme analizinin daha önce kullanıldığı çalışma alanlarına ve çalışmalara yer verilmiştir.

Dördüncü bölümde, çalışmada kullanılan Denizli kenti için 2004, 2005 ve 2006 yıllarına ait trafik kaza verilerine ve kümeleme analizinde kullanılacak parametrelere yer verilmiştir.

Beşinci bölümde, kümeleme analizinin yapıldığı bilgisayar programlarından kısaca bahsedilmiş ve trafik kaza verileri, bu bilgisayar programları ile kümeleme analizine tabi tutulmuş, elde edilen sonuçlar tablo ve şekillerle ortaya konmuştur.

Altıncı bölümde, elde edilen sonuçlar değerlendirilerek, ileriye yönelik yapılması gereken çalışmalara ilişkin öneriler sunulmuştur.

2. KÜMELEME ANALİZİ

2.1 Genel

Kümeleme analizi, bir araştırmada incelenen birimleri aralarındaki benzerliklerine göre belirli gruplar içinde toplayarak sınıflandırma yapmayı, birimlerin ortak özelliklerini ortaya koymayı ve bu sınıflar ile ilgili genel tanımlamalar yapmayı sağlayan bir yöntemdir (Şahin ve Hamarat 2002). Analiz sonucu elde edilen kümeler yüksek düzeyde küme içi homojenlik ve yüksek düzeyde kümeler arası heterojenlik gösterirler (Sharma 1996).

Kümeleme analizi, kümelerin sayısına veya küme yapılarına ilişkin herhangi bir varsayımda bulunmaz. Diğer çok değişkenli istatistiksel analiz yöntemlerinde önemli bir yer tutan normallik varsayımı, bu analizde prensipte kalmakta ve uzaklık değerlerinin normalliği yeterli görülmektedir.

Kümeleme Analizi X veri matrisinde yer alan ve doğal gruplamaları kesin olarak bilinmeyen birimleri, değişkenleri ya da birim ve değişkenleri birbirleri ile benzer olan alt kümelere ayırmaya yardımcı olan yöntemler topluluğudur. Kümeleme analizi; birimleri p değişkene göre hesaplanan ve benzerlik ölçüsü olarak kullanılan bazı ölçüler kullanarak homojen gruplara bölmek amacıyla kullanılır. Bu amaçlar dört grupta toplanabilir;

1) n sayıda birimi, nesneyi, oluşumu, p değişkene göre saptanan özelliklerine göre olabildiğince kendi içinde türdeş (homojen) ve kendi aralarında farklı (heterojen) alt gruplara (küme) ayırmak.

2) p sayıda değişkeni, n sayıda birimde saptanan değerlere göre ortak özellikleri açıkladığı varsayılan alt kümelere ayırmak ve ortak faktör yapıları ortaya koymak.

3) Hem birimleri hem değişkenleri birlikte ele alarak, ortak n birimi p değişkene göre ortak özellikli alt kümelere ayırmak.

4) Birimleri, p değişkene göre saptanan değerler için, izledikleri biyolojik ve tipolojik sınıflamayı ortaya koymak (taksonomik sınıflandırma yapmak).

Kümeleme işleminin uygulandığı veri setindeki her bir veriye nesne adı verilir ve kümeleme analizi, bu nesnelere benzerlik esasına göre birbirine benzeyen nesnelere aynı kümelerde toplar. Bu nesnelere iki boyutlu düzlem üzerinde noktalarla gösterilir. Benzerlikler, nesne çiftleri arasındaki uzaklığın ölçüsüdür. X veri matrisinde yer alan n birimin p değişkene göre uzaklıkları, uzaklık matrisi adı verilen D matrisi ile gösterilir. D matrisinin elemanları dij biçiminde gösterilir. Birimlerin birbirleri ile olan benzerlik düzeyleri, benzerlik matrisi Sim, elemanları da simij biçiminde gösterilir. Birimlerin benzerlikleri $simij = 100 (1 - dij / maxdij)$ biçiminde hesaplanır.

Kümeleme analizinin uygulama aşamaları;

Veri matrisinin belirlenmesi: Birim ya da değişkenlerin doğal gruplamaları hakkında kesin bilgilerin bulunmadığı popülasyonlardan alınan n sayıda birimin p sayıda değişkene ilişkin gözlemlerin elde edilmesidir.

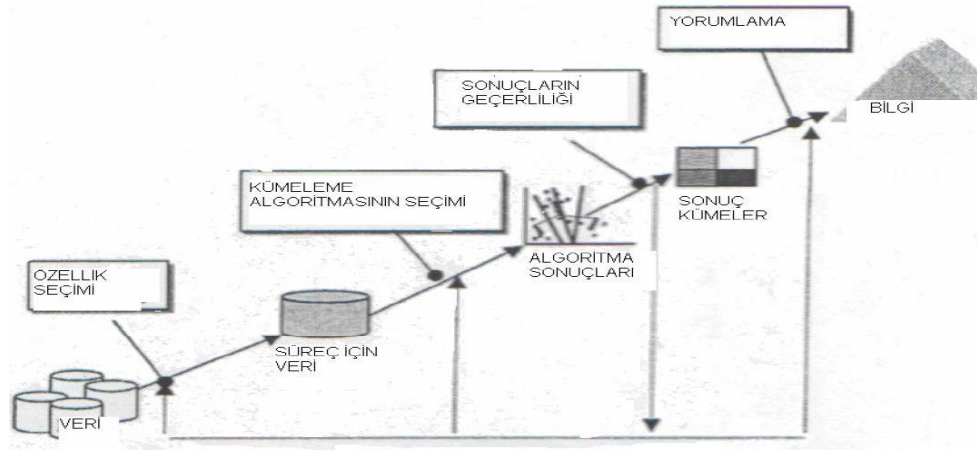
Benzerlik ya da farklılık matrisinin belirlenmesi: Birimlerin/değişkenlerin birbirleri ile olan benzerliklerini ya da farklılıklarını gösteren uygun bir benzerlik ölçüsü ile birimlerin/değişkenlerin birbirlerine uzaklıklarının hesaplanmasıdır.

Kümelere ayırma: Uygun kümeleme yöntemi yardımı ile benzerlik/farklılık matrislerine göre birimlerin/değişkenlerin uygun sayıda kümelere ayrılmasıdır.

Yorumlama: Elde edilen kümelerin yorumlanması ve bu kümeleme yapısına dayalı olarak kurulan hipotezlerin doğrulanması için gerekli analitik yöntemlerin uygulanmasıdır.

Yaygın olarak kullanılan kümeleme yöntemleri birimler arasındaki uzaklıklara dayanan benzerlik ya da benzemezlik matrisine göre işlem yaptıklarından, farklı kümeleme yöntemleri farklı uzaklık ölçülerine göre farklı sonuçlar verebilmektedir. Ayırmaya dayanan kümeleme yöntemleri her veri setinin her bir birimini bir ve yalnızca bir kümeye ayırır. Böylelikle aşamalı ya da aşamalı olmayan kümeleme yöntemleri her bir birim için kesin karar alırlar ve bir kümeye atarlar. Sonuçları itibarıyla yaklaşık aynı sonuçları veren kümeleme algoritmalarında bazı birimlerin farklı kümelerde yer aldığı

gözlenebilmektedir (Hamarat 1998). Bu tip durumlarda birimlerin küme üyeliklerinde bir bulanıklık söz konusu olmaktadır. Bir diğer ifadeyle birimlerin küme üyeliklerinde bir kararsızlık ortaya çıkmaktadır. Bulanık Kümeleme (Fuzzy Clustering) Yöntemi bu tip durumları tanımlamak için daha iyi bir yöntemdir.



Şekil 2.1 Kümeleme sürecinin adımları (Güler 2006)

Şekil 2.1’ de de görüldüğü üzere kümeleme sürecinin adımları aşağıdaki gibidir;

Özellik Seçimi: Amaç, ilgilenilen konuda mümkün olduğu kadar çok bilgiyi kodlayabilen, kümeleme ile ilgili özellikleri doğru dürüst bir şekilde seçmektir. Bu yüzden, verilerin kümeleme adımlarından önce işlenmesi gerekli olabilir.

Kümeleme Algoritması: Bu adım, veri seti için iyi bir kümeleme tasarımının tanımından ortaya çıkan algoritmanın seçimiyle ilgilidir. Yakınlık ölçüsü ve kümeleme kriteri çoğunlukla, veri setinin yapısına uygun kümeleme tasarımını tanımlamak için oldukça hızlı ve verimli çalışan kümeleme algoritmasını karakterize eder.

Sonuçların Geçerliliği: Kümeleme algoritmasının sonuçlarının doğru olup olmadığı uygun kriter ve tekniklerle test edilebilir. Kümeleme algoritmaları önceliği bilinmeyen kümeleri tanımladığından, kümeleme metotlarına bakılmaksızın, verinin sonuç bölünmesi çoğu uygulamada bazı değerlendirmeler gerektirir.

Sonuçların Yorumu: Birçok durumda, uygulama alanındaki uzman kişiler doğru karara varmak için diğer deneysel kanıtları da göz önüne alarak küme sonuçlarını değerlendirmek zorundadır.

2.2 Kümeleme Metotları

— Bölümlemeli Kümeleme Metodu

* Klasik Bölümlendirme Metodu

- k – ortalamalar
- k – medoids

* Modern Bölümlendirme Metodu

- Hard k - ortalamalar
- Bulanık c – ortalamalar

— Hiyerarşik Kümeleme Metodu

* Toplayıcı Metot

- Tek Bağlantılı Kümeleme
- Tam Bağlantılı Kümeleme
- Ortalama Bağlantılı Kümeleme
- Ortalama Grup Bağı
- Ward Yöntemi

* Bölen Metot

- DIANA

— Yoğunluk Tabanlı Kümeleme Metodu

* DBSCAN

* OPTICS (Ordering Points to Identify the Clustering Structure)

* FJP (Fuzzy Joint Points)

— Grid Bazlı Kümeleme Metodu

* STING

* WaveCluster

* CLIQUE

2.2.1 Bölümlenmeli kümeleme metodu

Bölümlenmeli kümeleme metodunda bireyler, başlangıç bölünmesinden başlanarak, yinelemeli olarak kümeleme kriterini azaltacak şekilde kümelere tahsis edilir.

Bölümlenmeli kümeleme yapılabilmesi için gerekli kriterler;

- Başlangıç Koşulları
- Küme Temsil Şeması
- Tahsis Etme Fonksiyonları
- En İyilik Kriteri
- Birleştir – Ayr Koşulları

Bir küme temsili, genellikle küme içerisindeki bireyleri karakterize eden matematiksel veya geometrik bir yapıdır. Mümkün temsil şemaları arasında; kümenin merkezi, kümenin en uçtaki üç bireyi, normal dağılım fonksiyonu, sınıflandırma ağacı, bağlayıcı ifadeler yer alır. Verilen bir küme temsil şeması, küme temsil fonksiyonu, verilen küme için en iyi temsili belirler.

Küme en iyilik kriteri, kümelerin doğruluğunu belirleyen bir ölçüdür. Bu kriter ayrıca yinelemenin devam edip etmeyeceğini belirtir. Küme en iyilik kriteri bu kümeleme algoritması için oldukça önemlidir. Doğru seçilmediği durumda yanlış kümelerin oluşmasına sebep olacaktır.

2.2.1.1 Klasik bölümlendirme metodu

Küme temsilcilerinin nasıl oluşturulacağına bağlı olarak, klasik bölümlendirme metodu k-ortalamar ve k-medoids olmak iki gruba ayrılır.

2.2.1.1.1 K-ortalamar metodu

K-ortalamar metodu, endüstriyel ve bilimsel anlamda kullanılan en popüler kümeleme metotlarından birisidir. Adı, kitle merkezi (centroid) olarak adlandırılan, k tane kümenin her birini temsil eden noktaların ortalamasından gelir. K-ortalamar kümeleme analizi öklit uzaklığını kullanır. Araştırmacı arzulanan küme sayısını baştan belirlemek zorundadır. K-ortalamar algoritması aşağıdaki amaç fonksiyonunu minimize etmeye çalışır;

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (2.1)$$

Burada, $\|x_i^{(j)} - c_j\|^2$, $x_i^{(j)}$ verisi ile c_j küme merkezi arasındaki mesafedir. J , n verinin kendi küme merkezlerinden olan toplam uzaklığıdır.

K-ortalamalar metodu, ilk önce n adet nesneden k adet nesne seçer ve bu nesnelerin her biri, bir kümenin merkezini veya orta noktasını temsil eder. Geriye kalan nesnelere her biri kendisine en yakın olan küme merkezine göre kümelere dağılırlar. Ardından her küme için ortalama hesaplanır ve hesaplanan bu değer o kümenin yeni merkezi olur. Bu işlem tüm nesnelere kümelere yerleşinceye kadar devam eder (Han ve Kamber 2000).

K-ortalamalar algoritması aşağıdaki adımlardan oluşur;

1. Başlangıç küme merkezleri olarak k nokta seç,
2. Her nesneyi kendine en yakın küme merkezi olan gruba ata,
3. Tüm nesnelere atandığında, k merkezin yerini yeniden hesapla,
4. 2. ve 3. adımları küme merkezleri sabitleninceye kadar tekrarla.

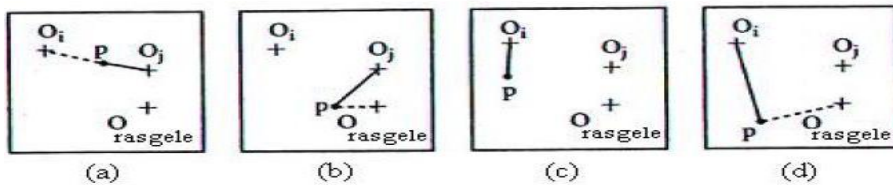
K-ortalamalar metodu, sadece kümenin ortalamasının tanımlanabildiği durumlarda kullanılabilir. Kullanıcıların k değerini, yani oluşacak küme sayısını belirtme gerekliliği bir dezavantaj olarak görülebilir. Esas önemli olan dezavantaj ise dışarıda kalanlar (outliers) olarak adlandırılan nesnelere karşı olan duyarlılıktır. Değeri çok büyük olan bir nesne, dahil olacağı kümenin ortalamasını ve merkez noktasını büyük bir derecede değiştirebilir. Bu değişiklik kümenin hassasiyetini bozabilir. Bu sorunu gidermek için kümedeki nesnelerin ortalamasını almak yerine, kümede ortaya en yakın noktada konumlanmış nesne anlamındaki medoid kullanılabilir.

2.2.1.1.2 K-medoidler metodu

K-medoidler kümeleme metodunun temel stratejisi ilk olarak n adet nesnede, merkezi temsili bir medoid olan k adet küme bulmaktır. Geriye kalan nesnelere, kendilerine en yakın olan medoide göre k adet kümeye yerleşirler. Bu bölünmelerin

ardından kümenin ortasına en yakın olan nesneyi bulmak için medoid, medoid olmayan her nesne ile yer değiştirir. Bu işlem en verimli medoid bulunana kadar devam eder.

Şekil 2.2’de O_i ve O_j iki ayrı kümenin medoidlerini, $O_{rasgele}$ seçilen ve medoid adayı olan bir nesneyi, p ise medoid olmayan bir nesneyi temsil etmektedir. Şekil 2.2 $O_{rasgele}$, şu anda şu anda medoid olan O_j ’nin yerine geçip, yeni medoid olup olmayacağını belirleyen dört durumu göstermektedir.



+ (kütle merkezi), _ (yer değiştirmeden önce), -- (Yer değiştirmeden sonra)

Şekil 2.2 k-medoidler metoduyla kümeleme örneği (Avcı 2005)

(a) : p nesnesi şu anda O_j medoidine bağlıdır (O_j medoidinin bulunduğu kümededir). Eğer O_j , $O_{rasgele}$ ile yer değiştirir ve p O_i ’ye en yakınsa, p nesnesi O_i ’ye geçer.

(b) : p nesnesi şu anda O_j medoidine bağlıdır. Eğer O_j , $O_{rasgele}$ ile yer değiştirir ve p $O_{rasgele}$ ’ye en yakınsa, p nesnesi $O_{rasgele}$ ’ye geçer.

(c) : p nesnesi şu anda O_i medoidine bağlıdır. Eğer O_j , $O_{rasgele}$ ile yer değiştirir ve p hala $O_{rasgele}$ ’ye en yakınsa, p nesnesi yine O_i ’ye bağlı kalır.

(d) : p nesnesi şu anda O_i medoidine bağlıdır. Eğer O_j , $O_{rasgele}$ ile yer değiştirir ve p $O_{rasgele}$ ’ye en yakınsa, p nesnesi $O_{rasgele}$ ’ye geçer.

K-medoidler algoritması aşağıdaki adımlardan oluşur;

1. k tane nesne seç (medoid),
2. Nesneleri onlara en yakın medoidlere at,
3. Bu nesne bir medoidmiş gibi ele alınıp toplam performans hesaplanır,
4. Eğer daha performanslı sonuç elde ediliyorsa diğeri yerine yeni medoid bu nesne olur (yer değiştirilir),
5. Bir değişiklik olmayana dek tekrarla.

k-medoidler, k-ortalamlar a göre sapan veriden daha az etkilenir.

2.2.1.2 Modern bölümlendirme metodu

2.2.1.2.1 Hard k-ortalamlar metodu

K-ortalamlar algoritması N vektörü c gruba ayırır ($G_i, i = 1, \dots, c$ kümelerdir). Bu algoritma, her bir nesnenin bir kümeye atandığı veya atanmadığı kesin kümeleme yapar. Algoritmanın amacı her grup için küme merkezlerini (centroid) bulmaktır. Algoritma (2.2) eşitliğinde verilen mesafe fonksiyonunu minimize eder;

$$J = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{k, x_k \in G_i} d(x_k - c_i) \quad (2.2)$$

c_i : i. kümenin merkezi, $d(x_k - c_i)$: i. merkez (c_i) ile k. veri noktası arasındaki mesafedir.

Basitleştirmek amacıyla, öklit mesafesi farklılık ölçüsü olarak kullanılmıştır ve farklılık fonksiyonu (2.3) eşitliğindeki gibi verilmiştir.

$$J = \sum_{i=1}^c J_i = \sum_{i=1}^c \left(\sum_{k, x_k \in G_i} \|x_k - c_i\|^2 \right) \quad (2.3)$$

Ayrılmış gruplar $c \cdot n$ boyutlu ikili üyelik matrisi U ile tanımlanabilirler. Bu matriste, j. veri noktası x_j i. gruba aitse $u_{ij} = 1$, aksi takdirde $u_{ij} = 0$ olarak alınır. Bunun formüle edilmiş şekli (2.4)'deki gibidir.

$$u_{ij} = \begin{cases} 1, & \text{Eger } \|x_j - c_i\|^2 \text{ ise, } \forall k \neq i \\ 0 & \text{aksitakdirde} \end{cases} \quad (2.4)$$

Bir nokta sadece bir grup içinde olabileceği için, üyelik matrisi U aşağıda gösterilen iki özelliğe sahiptir.

$$\sum_{i=1}^c u_{ij} = 1, \forall j = 1, \dots, n \quad (2.5)$$

$$\sum_{i=1}^c \sum_{j=1}^n u_{ij} = n \quad (2.6)$$

Merkezler, i . gruptaki tüm vektörlerin ortalaması olarak hesaplanır.

$$c_i = \frac{1}{|G_i|} \sum_{k, x_k \in G_i} x_k \quad (2.7)$$

$|G_i|$, G_i 'nin boyutudur.

$x_j, j = 1, \dots, n$ veri seti için k-ortalamlar algoritması,

1. $c_i = i = 1, \dots, c$ merkezlerine ilk değerleri atanır. Bu temelde, tüm veri noktalarından rasgele olarak c adet nokta seçerek başılır.

2. (2.4) eşitliği ile U üyelik matrisi tanımlanır.

3. (2.3) eşitliği kullanılarak farklılık fonksiyonu hesaplanır. Önceki iterasyonun üzerinde artış eşikten küçükse dur.

4. (2.7) eşitliğini kullanarak yeni merkezleri hesapla ve 2. adıma dön.

Algoritmanın performansı merkezlerin başlangıç yerlerine bağlıdır. Yani algoritma optimal çözüm için garanti vermez.

2.2.1.2.2 Bulanık c-ortalamlar metodu

Klasik kümeleme metodunda, bir elemanın bir kümeye veya bir sınıfa ait olması, ya aittir (üyelik= 1) ya da ait değildir (üyelik= 0) şeklinde açıklanmaktadır. Gerçekte bir eleman bir kümeye ne tam aittir ne de değildir. Yani bu elemanın o küme için bir aittik derecesi (üyelik değeri) olmalıdır. Bu üyelik değeri 0 ile 1 arasında sonsuz değer alabilmektedir ve bu üyelik değerlerinin toplamı daima 1'e eşittir. Böylelikle eleman **en yüksek üyelik katsayısına** sahip olduğu kümeye atanır. Birbirine çok benzeyen birimler aynı kümede yüksek üyelik ilişkisine göre yer alırlar. Bulanık algoritmalarda,

bir eleman kümelendirirken veya sınıflandırılırken elemanın sınıfını belirlemenin yanında o sınıfa ne kadar ait olduğuna dair bir bilgi de verildiğinden bulanık algoritmalar, klasik algoritmalara oranla daha fazla bilgi içerirler (Höppner vd. 1999). Bu tip algoritmalar gerçek sayıların belirsizliğini ele aldığından günlük yaşamın tecrübelerine uygun kümeleme şekillerinin ortaya çıkmasını sağlar.

Bulanık c-ortalamlar kümeleme metodu, kesin bölümlenmeyi kullanan hard k-ortalamlar metodundan bulanıklık yönünden farklılaşır. Bulanık c-ortalamlar, veri noktasının tüm gruplara 0 ile 1 arasında üyelik dereceleri ile ait olacak şekilde bulanık bölümlenmesini gerçekleştirir.

Bulanık c-ortalamlar iteratif bir algoritmadır ve amacı bir farklılık fonksiyonunu minimize eden küme merkezlerini (centroid) bulmaktır.

Bulanık kümeleme metodunda, ilk olarak (2.8) eşitliğine göre U üyelik matrisine rasgele değer atanır.

$$\sum_{i=1}^c u_{ij} = 1, \forall_j = 1, \dots, n \quad (2.8)$$

Bulanık c-ortalamlar metodunda kullanılan farklılık fonksiyonu,

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 \quad (2.9)$$

u_{ij} : 0 ile 1 arasındadır,

c_i : i. kümenin merkezi,

d_{ij} : i. merkez c_i ile j. veri noktası arasındaki öklit mesafesi,

$m \in [1, \infty]$: ağırlıklandırma faktörü.

Farklılık fonksiyonunu minimum yapmak için iki koşul vardır. Bu koşullarda (2.10) ve (2.11) eşitliklerinde verilmiştir.

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (2.10)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad (2.11)$$

Bulanık c-ortalamalar algoritması Bezdek (1973) tarafından şu şekilde verilmiştir;

1. (2.8) denklemini sağlayacak şekilde U üyelik matrisine rasgele değerler atanır.
2. (2.10) denklemini kullanılarak c_i merkezleri hesapla,
3. (2.9) denklemini kullanarak merkezler ile veri noktaları arasındaki farklılığı hesapla. Önceki iterasyon üzerine artışı eşik değerin altındaysa dur,
4. (2.11) denklemini kullanarak yeni bir U matrisi hesapla. Adım 2'ye git.

Bulanık c-ortalamalar, küme merkezlerini ve her veri noktası için üyelik derecelerini iteratif olarak güncellemekle, küme merkezlerini veri seti içinde doğru yere taşır.

Küme merkezlerinin ilk yerleri, başlangıçta değeri rasgele atanan U matrisi kullanılarak oluşturulduğu için, bulanık c-ortalamalar optimal sonuca yaklaşmayı garanti etmez (Sintas vd. 1999).

Performans merkezlerin başlangıç küme merkezlerine bağlıdır. Daha güçlü bir yaklaşım için aşağıda tanımlanan iki yol vardır.

1. Tüm merkezleri tanımlamak için bir algoritma kullanmak (örneğin: tüm veri noktalarının aritmetik ortalaması)

2. Bulanık c-ortalamları farklı başlangıç merkezleri ile tekrarlı olarak çalıştırmak.

2.2.1.2.2.1 Fanny algoritması

Bu algorithmanda kullanılan bulanık kümeleme tekniği aşağıdaki amaç fonksiyonunun minimizasyonunu amaçlar. Bu amaç fonksiyonunda üyelik fonksiyonları şu kısıtlara sahiptir:

$$1 - u_{iv} \geq 0 \text{ ise } i=1, \dots, n \text{ ve } v=1, \dots, k$$

$$2 - \sum_{v=1}^k u_{iv} = 1 = \%100 \text{ ise } i=1, \dots, n$$

Burada her bir i birimi ve her bir v kümesi u_{iv} 'nin bir üyesi olacaktır. u_{iv} , i . birimin v kümesine ne kadar ait olduğunu gösterir. Bu şartlar altında amaç fonksiyonu (2.12) eşitliğindeki gibidir.

$$C = \sum_{v=1}^k \frac{\sum_{i,j=1}^n u_{iv}^2 u_{jv}^2 d(ij)}{2 \sum_{j=1}^n u_{jv}^2} \quad (2.12)$$

Burada, $d(i, j)$, i ve j . birimler arasındaki uzaklık (benzerlik), u_{iv} ise i . birimin v . kümeye bilinmeyen üyeliğini tanımlar. Bulanık kümelemede her bir birimin tüm kümelere olan üyelik katsayıları toplamı daima 1 olacak şekilde pozitifdir. Bulanık kümelemenin, kesin kümelemeden ne kadar uzakta olduğu Dunn ayrıştırma katsayısıyla değerlendirilir. Bu katsayı elde edilen kümenin ne kadar bulanık olduğuna ilişkin bir fikir verir. Dunn Ayrıştırma Katsayısı, tüm üyelik katsayılarının kareler toplamının birim sayısına bölünmesiyle eşitlik (2.13)' deki gibi hesaplanır.

$$F_k = \sum_{i=1}^n \sum_{v=1}^k u_{iv}^2 / n \quad (2.13)$$

F_k her zaman $[1/k, 1]$ aralığında bulunur. Böylelikle birimlerin üyelikler matrisi elde edilir. Burada U aşağıdaki tanımlanan tüm üyelikler matrisidir ve her zaman $[1/k, 1]$ aralığında bulunur.

$$U = \begin{matrix} & \begin{matrix} 1 & \dots & k \end{matrix} \\ \begin{matrix} 1 \\ \vdots \\ k \end{matrix} & \begin{bmatrix} & & \\ & U_{iv} & \\ & & \end{bmatrix} \end{matrix}$$

Dunn Ayrıştırma Katsayısı;

1 - Tamamen bulanık kümeleme durumunda tüm $U_{iv} = \frac{1}{k}$ ise $F_k = nk \frac{1}{nk^2} = \frac{1}{k}$ olur.

2 - Kesin kümeleme durumunda tüm $U_{iv} = 0$ veya $F_k = \frac{n}{n} = 1$ olur. Dolayısıyla

Dunn katsayısı 0 durumunu tamamen bulanık ve 1 durumunu da kesin küme durumları olarak göstermektedir. Küme sayısından bağımsız olarak 1'den (kesin kümeden) 0'a (bütünüyle bulanık) eşitlik göstermek üzere, bu katsayı küme sayısından bağımsız olarak normalleştirilirse,

$$F_k^1(u) = \frac{F_k(u) - \frac{1}{k}}{1 - \frac{1}{k}} = \frac{kF_k(u) - 1}{k - 1} \quad (2.14)$$

şeklinde normalleştirilebilir. Bu şekilde Normalleştirilmiş Dunn Katsayısı elde edilir, bu katsayı $[0-1]$ aralığında yer alır ve Bulanıksızlık Endeksi (Norfuzziness Index) olarak isimlendirilir.

Bulanık kümeleme yönteminde birimlerin ne derece iyi kümelendikleri Siluet Katsayısı (silhouette coefficient), $s(i)$ ile belirlenir. $s(i)$, i . birimin ne derece iyi kümelendiğini gösterir ve $-1 \leq s(i) \leq 1$ arasında yer alır. $s(i)$ bir kümedeki tüm birimler için Ortalama Siluet Görüntü Katsayısıdır. $s(i)$ de; tüm birimler için k küme sayısına göre birimlerin ne derece iyi kümelendiklerini gösterir katsayıdır ve $s(i)$ 'lerin

ortalamasına eşittir. En büyük $s(i)$ 'e karşılık gelen küme sayısı k , en uygun küme sayısı olarak seçilir (Bezdek ve Pal 1992)

2.2.1.2.2.2 Doğrulama

Bulanık kümelemedeki ana problemlerden birisi küme sayısının baştan belirlenmesi gerekliliğidir. Farklı başlangıç küme sayısı seçimleri, farklı kümelemelerin ortaya çıkmasına sebep olur. Onun için, kümeleme analizinden sonra her bir bulanık bölmenin doğrulamasının yapılması gereklidir. Kümeleme algoritmaları her zaman belirlenmiş küme sayıları için en uygun formu bulmaya çalışır. Her ne kadar durum böyle olsa da, bu en uygun formun anlamlı olduğu anlamına gelmez. Küme sayısı yanlış olabilir veya küme şekilleri verideki gruplara uymayabilir (Murat vd 2008). Bu çalışmada birkaç kümeleme indeksi kullanılarak optimum başlangıç küme sayısı tespit edilmiş ve bu küme sayısına göre veri seti bulanık kümelemeye tabi tutulmuştur. Bu indeksler;

- Ayırma Katsayısı (Partition Coefficient, PC)
- Sınıflandırma Entropisi (Classification Entropy, CE)
- Ayırma İndeksi (Partition Index, SC)
- Ayırıştırma İndeksi (Separation Index, S)
- Xie ve Beni İndeksi (Xie and Beni's Index, XB)
- Dunn İndeksi (Dunn's Index, DI)

Ayırma katsayısı (PC), iki bulanık kümenin üst üste gelme miktarını ölçer. Verilerin özelliklerine direk bağlantısının eksik olması bu indeksin dezavantajıdır. Optimum küme sayısı, bu indeksin maksimum değerine karşı gelen küme sayısıdır. İndeks aralığı $[1/c, 1]$.

$$PC(c) = \frac{1}{N} \sum_{i=1}^c \sum_{k=1}^N (u_{ik})^2 \quad (2.15)$$

Sınıflandırma entropisi (CE), küme ayrımlarının bulanıklığını ölçer. Optimum küme sayısı, bu indeksin minimum değerine karşı gelen küme sayısıdır. İndeks aralığı $[0, \log a(c)]$.

$$CE(c) = -\frac{1}{N} \sum_{i=1}^c \sum_{k=1}^N u_{ik} \log_a a u_{ik} \quad (2.16)$$

Ayırma indeksi (SC), toplam sıklık ve küme ayırma oranıdır. SC kullanışlı bir indekstir ve bu indeksin düşük değeri, daha iyi bir ayırmanın ispatıdır.

$$SC(c) = \sum_{i=1}^c \frac{\sum_{k=1}^N (u_{ik})^m \|x_k - v_i\|^2}{N_i \sum_{k=1}^c (u_{ik})^m \|v_k - v_i\|^2} \quad (2.17)$$

Ayrışma indeksi (S), geçerli bölünme için minimum uzaklık ayrışmasını kullanır.

$$S(c) = \frac{\sum_{i=1}^c \sum_{k=1}^N (u_{ik})^2 \|x_k - v_i\|^2}{N \min_{i,k} \|v_k - v_i\|^2} \quad (2.18)$$

Xie ve Beni İndeksi (XB), kümeler arasındaki toplam varyasyon oranının ve kümelerin ayrışma oranının miktarını belirlemeyi amaçlar. Optimum küme sayısı, bu indeksin minimum değerine karşı gelen küme sayısıdır.

$$XB(c) = \frac{\sum_{i=1}^c \sum_{k=1}^N (u_{ik})^m \|x_k - v_i\|^2}{N \min_{i,k} \|x_k - v_i\|^2} \quad (2.19)$$

Dunn indeksi (DI), sıklık teşhisini ve ayrık küme kullanımını önerir.

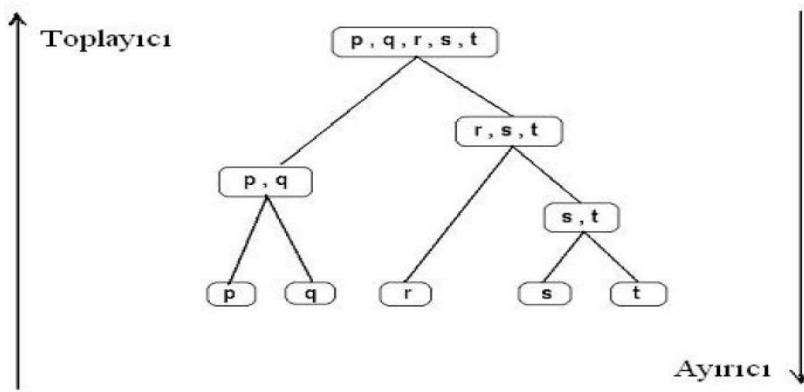
$$DI(c) = \min_{i \in C} \left\{ \min_{k \in C, i \neq k} \left\{ \frac{\min_{x \in C_i, y \in C_k} d(x, y)}{\max_{k \in C} \left\{ \max_{x, y \in C} d(x, y) \right\}} \right\} \right\} \quad (2.20)$$

2.2.2 Hiyerarşik (aşamalı) kümeleme metodu

Bölümlendirme algoritmalarının tüm kümeleri bir anda tanımlamasının aksine hiyerarşik algoritmalar önceden bulunan kümeleri kullanarak izleyen kümeleri bulur. Hiyerarşik kümeleme aşamalı olarak daha küçük kümelerin daha büyük kümelere birleşmesi (toplayıcı: alttan – üste) veya büyük kümelerin daha küçük kümelere ayrılmasıdır (bölücü: üstten – alta). Hiyerarşik kümeleme metodu *dendogram* (Şekil

2.3'de gösterildiği üzere) olarak bilinen, kümelerin ağaç yapısını veya diğer bir deyişle bir küme hiyerarşisini ortaya çıkarır. Her küme düğümü, ortak ebeveynlere sahip noktaların çocuk ve kardeş küme bölünmelerini içerir. Böyle bir yaklaşım, farklı seviyelerdeki düğümlerin dikkatle incelenmesine olanak sağlar.

Hiyerarşik kümelemede veriler kümelere tek bir adımda bölünemezler. Onu yerine, tüm nesnelere içeren tek bir kümeden, her biri tek nesne içeren n kümeye doğru bir seri bölünme uygulanır.



Şekil 2.3 Hiyerarşik kümeleme için dendrogram

Hiyerarşik kümelemenin *avantajları*:

- Düğümlerin seviyesine ilişkin esneklik özelliğinin olması
- Benzerlik ve uzaklık biçimlerini ele almanın kolaylığı
- Çeşitli niteliklere uygulanabilme özelliğinin olması

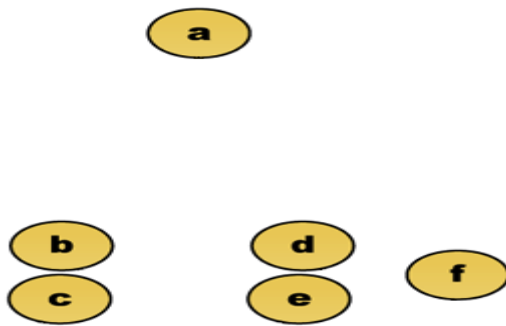
dezavantajları:

- Bitirme kriterinin belirsizliği
- Gerçekte çoğu hiyerarşik kümeleme algoritması, geliştirme amaçlarına göre arada bulunan kümelere (dendrogramda) tekrar ulaşılmasına izin vermez

olarak sıralanabilir.

2.2.2.1 Toplayıcı (agglomerative) hiyerarşik kümeleme metodu

Toplayıcı hiyerarşik kümeleme metodunda, kümeleme sürecinin başlangıcında her birey bir kümedir, süreç sonunda ise her bir birey tek bir kümede toplanır (Tatlıldil 2002). Toplayıcı hiyerarşik kümeleme metodu, verilerin bir seri işlem ile birleşimini oluşturur (P_n, P_{n-1}, \dots, P_1). İlk P_n , n tane nesneden (kümeden), son P_1 , n nesnenin tümünü içeren tek bir gruptan oluşur. Metot, her adımda birbirlerine en yakın iki kümeyi bir araya getirir.

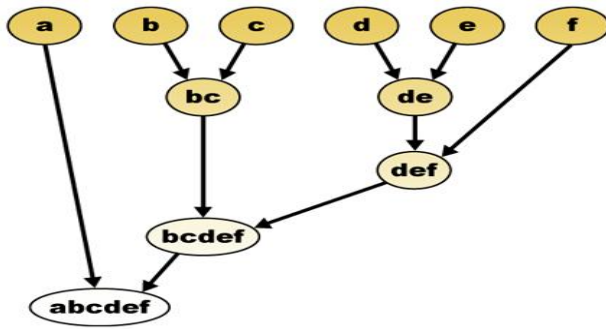


Şekil 2.4 Ham veri (WEB_1: http://en.wikipedia.org/wiki/Data_clustering)

Toplanmış hiyerarşik kümeleme metoduna ait algoritma şu şekilde olabilir;

1. n tane birey, n tane küme olmak üzere işleme başlanır.
2. En yakın (d_{ij} değeri en küçük olan) iki küme birleştirilir.
3. Küme sayısı bir azaltılarak yinelenmiş uzaklıklar matrisi bulunur.
4. 2 ve 3 nolu adımlar n-1 kez tekrarlanır (Tatlıldil 2002).

Hiyerarşik kümeleme dendogramı şu şekilde olabilirdi;



Şekil 2.5 Geleneksel temsili

Kümeler arasında mesafeyi (benzerliği) tanımlamanın çeşitli yolları olduğu için metotlar arasında farklılıklar ortaya çıkar. Bu metotlar sırasıyla tek bağlantılı, tam bağlantılı, grup ortalama, merkezi, ortanca, minimum varyans ve Ward teknikleridir.

2.2.2.1.1 Tek bağlantılı kümeleme yöntemi (TekBKY)

En Yakın Komşuluk (Nearest Neighbour-Single Linkage-SLINK) olarak da bilinen bu teknikte uzaklıklar matrisi kullanılarak birbirine en yakın birey ya da kümeler birleştirilmekte ve birleştirme ardı ardına tekrarlanarak sürdürülmektedir (Tatlidil 2002).

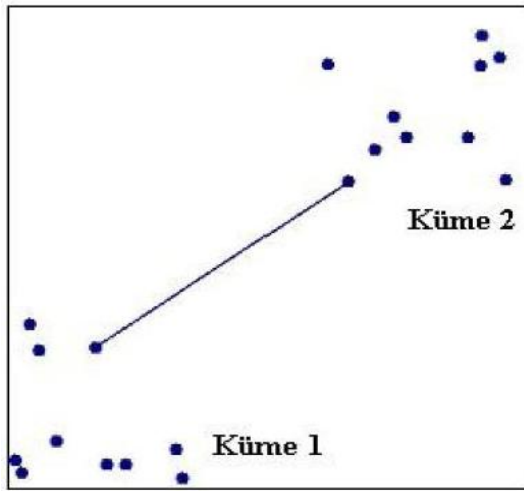
Tek bağlantılı kümeleme metodunda iki küme arasındaki uzaklık ölçütü $D(r,s)$ aşağıdaki gibi hesaplanır.

$$D(r,s) = \text{Min}\{d(i,j) : i, r \text{ kümesin deki}, j, s \text{ kümesin deki nesnedir}\}$$

Burada olası tüm (i,j) nesne çiftleri arasındaki mesafeler hesaplanır (i, r kümesindeki, j, s kümesindeki nesne). Bu mesafelerin en küçüğü r ve s kümeleri arasındaki mesafe olarak alınır. Diğer bir değişle, iki küme arasındaki mesafe, kümeler arası en kısa bağın değeri ile verilir.

Hiyerarşik kümelemenin her adımında, $D(r,s)$ değeri en küçük olan r ve s kümeleri birleştirilir.

Bu gruplar arası mesafe ölçüsü aşağıdaki Şekil 2.6' da gösterilmiştir.



Şekil 2.6 Tek bağlantılı kümeleme modeli

TekBKY ile hiyerarşik kümeleme yapmak için aşağıdaki işlem sırası izlenir:

1. X veri matrisinin D öklid uzaklık matrisi hesaplanır.

2. Eğer istenirse D matrisinden Sim (benzerlik) matrisi hesaplanır.

3. D ya da Sim matrisinde en küçük değerli birimler hiyerarşik olarak birbirleri ile birleştirilir. Küme (ij) oluşturulur. i. ve j. kümeleri birbirleri ile birleştirildikten sonra D (ya da Sim) matrisinde j. kümeye ilişkin satır silinir ve (ij) kümesinin uzaklığı (ya da benzerliği) i. kümenin uzaklığı olarak kalır ($d_i \leq d_j$).

4. Tüm kümeler birbirleri ile birleştirilinceğe kadar 3. adımdaki işlemler tekrarlanır (Özdamar 1997).

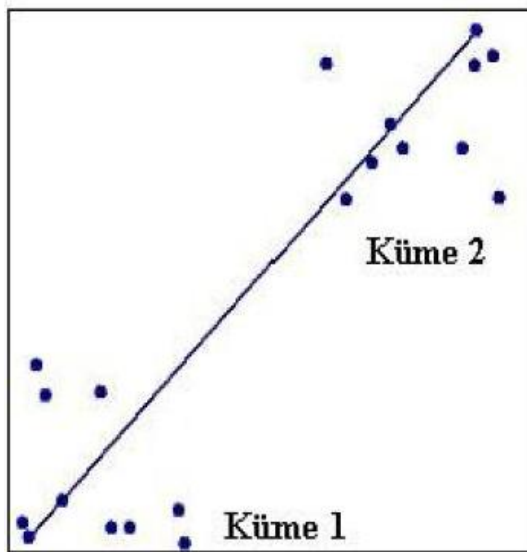
2.2.2.1.2 Tam bağlantılı kümeleme yöntemi (TBKY)

En Uzak Komşuluk (Furthest Neighbour Complete Linkage-CLINK) olarak da bilinen bu teknik yine Johnson tarafından önerilmiştir. Tek bağlantılı metodun tam tersi olan bu teknikte iki küme arasındaki uzaklık olarak, iki kümedeki eleman çiftleri arasındaki uzaklığın en büyüğü alınmaktadır.

$$D(r,s) = \text{Max}\{d(i,j):i,r \text{ kümesin deki}, j,s \text{ kümesin deki nesnedir}\}$$

Burada olası tüm (i,j) nesne çiftleri arasındaki mesafeler hesaplanır (i, r kümesindeki, j, s kümesindeki nesne). Bu mesafelerin en büyüğü r ve s kümeleri arasındaki mesafe olarak alınır. Diğer bir deyişle, iki küme arasındaki mesafe, kümeler arası en uzun bağıın değeri ile verilir.

Bu gruplar arası mesafe ölçüsü aşağıdaki Şekil 2.7' de gösterilmiştir.



Şekil 2.7 Tam bağlantılı kümeleme modeli

Tam bağlantılı kümeleme algoritması,

1. Her bireyi kendi kümesine yerleştir. Tüm sırasız birey çiftleri için bireyler arasındaki uzaklığın listesini oluştur ve listeyi küçükten büyüğe doğru sırala,
2. Birbirlerine en uzak (uzaklık değerleri en büyük) olan birey ya da kümeleri birleştir.
3. Bütün bireyler birleştirilen bir kümenin üyesi ise işlemi bitir. Aksi takdirde 2. adıma geri dön.

2.2.2.1.3 Ortalama bağlantılı kümeleme yöntemi (OrtBKY)

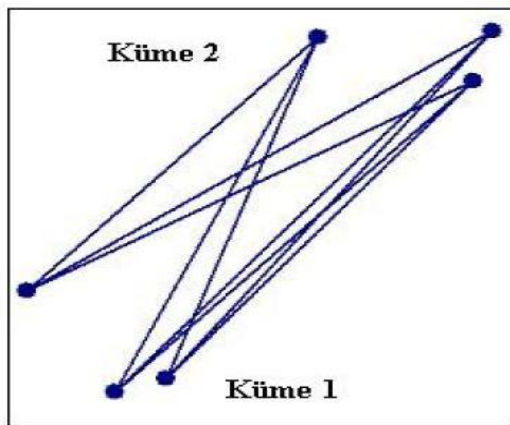
Tek bağlantılı teknikte işlemlerin uzun sürmesi, tam bağlantılı teknikte ise, aynı küme içersindeki bireylerin uzaklıklarının belli bir değerden küçük olması durumunda tüm kümelerin sağlıklı oluşturulmasının garanti edilememesi, son yıllarda sıkça kullanılan OrtBKY yönteminin alternatif olarak önerilmesine sebep olmuştur. Burada iki küme arası mesafe, her biri bir gruptan olacak olan tüm nesne çiftleri arasındaki ortalama mesafedir.

Ortalama bağlantılı kümeleme metodunda $D(r,s)$ aşağıdaki gibi hesaplanır.

$$D(r,s) = T_{rs} / (N_r * N_s)$$

T_{rs} : r ve s kümesi arasındaki tüm çiftlerin mesafeleri toplamıdır.

N_r ve N_s : r ve s kümelerinin boyutudur.



Şekil 2.8 Ortalama bağlantılı kümeleme modeli

2.2.2.1.4 Ortalama grup bağı

Bu metot ile oluşturulan gruplar, her değişken için ortalama değerleri ile tanımlanırlar. Ortalama vektörü ve gruplar arası mesafe bu iki ortalama vektörü arasındaki mesafe açısından tanımlanır.

Ortalama grup bağı metodunda r ve s kümeleri, birleştirmeden sonra yeni oluşturulan kümeler arası ortalama çiftler mesafesi minimum olacak şekilde birleştirilir. r ve s kümelerinin birleştirilmesiyle oluşan yeni kümeye t dersek, r ve s kümeleri arası mesafe $D(r,s)$ aşağıdaki gibi hesaplanır.

$D(r,s) = \text{Ortalama } \{ d(i,j) : i \text{ ve } j \text{ nesnelere } t \text{ kümesindedir, yeni küme } r \text{ ve } s \text{ kümelerinin birleştirilmesiyle oluşturulur} \}.$

Hiyerarşik kümelemenin her adımında, $D(r,s)$ değeri en küçük olan r ve s kümeleri birleştirilir. Bu durumda bu iki küme, yeni oluşturulan kümedeki ortalama noktalar arası çiftler mesafesi minimum olacak şekilde birleştirilir.

2.2.2.1.5 Ward'ın hiyerarşik kümeleme metodu

Bu metod kümeler arası mesafeyi hesaplamak için varyans analizi kullandığından diğer tüm yöntemlerden farklıdır. Bu metod, her adımda oluşturulabilecek herhangi iki kümenin kareler toplamını minimum yapmaya çalışır (Ward 1963). Genelde Ward'ın yöntemi oldukça etkilidir ancak küçük ölçülü kümeler yaratma eğilimindedir.

Ward, her grup ile ilgili kaybı minimize edecek şekilde P_n, P_{n-1}, \dots, P_1 bölümlenmelerini oluşturacak ve kolaylıkla yorumlanabilir formdaki kaydı ölçebilecek bir kümeleme işlemi önermiştir. Analizdeki her adımda, tüm mümkün küme çiftlerinin birleşimi ele alınır ve bilgi kaybında minimum artış yaşanan birleştirme sonucu görülen kümeler bir araya getirilir. Bilgi kaybı, hata kareler toplamı kriteri (error sum-of-squares criterion, ESS) ile tanımlanır.

Ward'ın önerisinin arkasındaki temel, en basit şekilde tek değişkenli veriler ele alınarak açıklanabilir. Örneğin (2, 6, 5, 6, 2, 2, 2, 0, 0, 0) değerlerine sahip 10 nesne olsun. 2,5 ortalama ile 10 değer tek bir grup olarak ele alındığında ortaya çıkan değer kaybı aşağıdaki ESS ile verilebilir.

$$ESS_{\text{BirGrup}} = (2-2,5)^2 + (6-2,5)^2 + \dots + (0-2,5)^2 = 50,5$$

Diğer yandan, 10 nesne değerlerine göre 4 küme ile sınıflandırılırsa ($\{0, 0, 0\}$, $\{2, 2, 2, 2\}$, $\{5\}$, $\{6, 6\}$), ESS 4 ayrı grubun hata kareler toplamının toplamı olarak hesaplanır.

$$ESS_{\text{BirGrup}} = ESS_{1.\text{Grup}} + ESS_{2.\text{Grup}} + ESS_{3.\text{Grup}} + ESS_{4.\text{Grup}} = 0$$

Böylece, 10 değeri 4 küme haline getirmek hiçbir bilgi kaybına neden olmaz.

2.2.2.2 Bölen metot

Bölen kümeleme metodu toplayıcı metodun tersidir ve başlangıçta tüm birimlerin bir küme oluşturduğunu kabul ederek birimleri aşamalı olarak n birimi sırasıyla 1, 2, 3, ..., n-1, n kümeye ayırmaya çalışan bir yaklaşımdır.

2.2.2.2.1 DIANA

Bölen bir hiyerarşik kümeleme tekniği olan DIANA' nın toplayıcı metottan ana farkı hiyerarşiyi ters sırada oluşturmasıdır. Sonuçlar hiyerarşik metot ile aynı değildir.

Toplayıcı metodun ilk adımında iki kümenin tüm mümkün birleşimleri için $[n(n-1)]/2$ kombinasyon vardır. Aynı prensibe bağlı bölen metotta ise verileri iki kümeye bölmek için $2^{n-1}-1$ olasılık söz konusudur. Bu olasılıkların sayısı toplayıcı metodundakiler ile karşılaştırıldığında önemli miktarda büyüktür.

Nesneler arası farklılıkları bulmak için korelasyon kullanılır. x ve y nesneleri arasındaki korelasyonu hesaplamak için korelasyon katsayısı kullanılır.

$$R(x, y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (2.21)$$

Korelasyon katsayısı $R(x,y)$ farklılık ölçüsünü temsil eder.

Korelasyonlardan farklılık matrisi oluşturulur ve diğer nesnelere ortalamada en farklı olan nesne ayırıcı grubu oluşturmak için seçilir. Ardından orijinal küme ile ayırıcı

küme arasındaki ortalama farklılıklar tekrar hesaplanır. Bu süreç, orijinal kümede kalan nesnelere birbirlerine ayırıcı gruptan daha fazla benzeyene kadar devam eder. Algoritma bir sonraki adımda, iki elemanı birbirinden en uzak olan kümeyi bulur ve kümeyi önceden bahsedilen şekilde ayırır. Kümeleme işlemi, tüm kümeler tek bir nesne içerene kadar devam eder.

Algoritma aşağıdaki gibidir,

1. Diğer tüm nesnelere karşı ortalama benzemezliği en yüksek olan nesne bulunur. Bu nesne ayırıcı *grup olarak* yeni bir küme yaratır.

2. Ayırıcı grubun dışındaki her i nesnesi için D_i mesafesi aşağıdaki gibi hesaplanır.

$$D_i = [\text{ortalama } d(i,j), j \notin \text{ayırıcı grup}] - [\text{ortalama } d(i,j), j \in \text{ayırıcı grup}]$$

3. D_h farkı en büyük olan h nesnesi bulunur. D_h pozitif ise, h ayırıcı gruba yakındır.

4. 2. ve 3. adımlar tüm D_h farkları negatif olana kadar tekrarlanır. Böylece veri seti iki kümeyle ayrılır.

5. En büyük çapa sahip küme seçilir. Bir kümenin çapı, kümenin herhangi iki nesnesi arasındaki en büyük farklılıktır. O zaman bu küme 1–4 adımları izlenerek bölünür.

6. Adımlar tüm kümeler tek nesne içerene kadar tekrar edilir.

2.2.2.2.1.1 Bölen katsayısı (BK)

Her i nesnesi için $d(i)$, nesnenin bölünmeden önce ait olduğu kümenin çapının, tüm veri setinin çapına oranıdır ve algoritma ile bulunan kümeleme yapısının gücünü gösterir.

$$BK = \frac{\sum d_i}{n} \quad (2.22)$$

2.2.3 Yoğunluk tabanlı kümeleme metodu

Yoğunluk tabanlı kümeleme, kümelere, veri seti içindeki bireylerin yoğun olduğu bölgeler gözüyle bakar ve bu veri setini düşük yoğunluktaki düşük yoğunluktaki bölgelere bölmeye çalışır. Bu kümelemeye göre noktaları sonlu sayıda kümeye bölmek için, yoğunluk, bağlanabilirlik ve sınır gibi kavramların bilinmesi gerekir. Yoğunluk tabanlı kümeleme algoritmaları keyfi şekilli kümeleri ortaya çıkarmada başarılıdır.

2.2.3.1 DBSCAN

DBSCAN isteğe bağlı şekilli kümeleri bulmak için tasarlanan yoğunluk tabanlı bir kümeleme algoritmasıdır. DBSCAN tek parametre gerektirir ve kullanıcıyı bunun uygun değerleri için destekler. Bu algoritmanın ana fikri, küme içindeki her nokta (bu noktalara merkez gözüyle bakıldığında) için; verilen bir yarıçap etrafında en az minimum sayıda nokta olmak zorundadır. Yani bir kümenin her noktası için verilen bir çapın komşuluğunun en az minimum sayıda nokta içermesidir, örneğin komşuluktaki yoğunluk bir eşik değerini aşmalıdır (Ester vd. 1996). Bu komşuluğun şekli $dist(p,q)$ ile gösterilen p ve q noktaları için mesafe fonksiyonunun seçimi ile tanımlanır.

DBSCAN uzaysal kümeleri bulmak için idealdir. Normal olarak her küme için Eps ve $MinPts$ değerleri ve kümeden herhangi bir nokta bilinmelidir. Bu şekilde doğru parametreler kullanılarak düzgün kümeler oluşturulabilir.

Kümeleri yoğunluk ile tanımlamamızın ana nedeni, her küme içinde kümenin dışından önemli miktarda büyük olan noktaların tipik bir yoğunluğu olmasıdır.

2.2.3.1.1 Bir noktanın eps _komşuluğu

Bir p noktasının Eps _Komşuluğu, $N_{Eps}(p)$ ile gösterilir ve $N_{Eps}(p) = \{q \in D \mid dist(p,q) \leq Eps\}$ ile tanımlanır. Kümedeki her nokta için, o noktanın Eps _Komşuluğundan en az $MinPts$ kadar minimum sayıda nokta olması yaklaşımı gereklidir. Ancak bu yaklaşım başarısız olur bir kümede iki tür nokta vardır, kümenin içindeki noktalar (core points) ve kümenin sınırındaki noktalar (border points). Genelde sınır noktasının Eps _Komşuluğu, öz noktanın Eps _Komşuluğu'ndan çok daha az nokta içerir. Bu yüzden, aynı kümeye ait olan tüm noktaları içermesi için minimum nokta sayısı göreceli olarak küçük seçilmelidir. Ancak bu değer sonraki kümeler için

karakteristik olmayacaktır. Böylece, bir C kümesindeki her p noktası için, p q 'nın $Eps_Komşuluğu$ 'nda ve $N_{Eps}(q)$ en az $MinPts$ kadar nokta içerecek şekilde bir C 'de q noktası olması önerilir.

2.2.3.1.2 Direkt yoğunluk ulaşılabilirlik

Bir q noktasının p noktasına direkt yoğunluk ulaşılabilir olması için Eps ve $MinPts$ koşulları altında şu özellikleri sağlaması gerekir.

- $p \in N_{Eps}(q)$
- $|N_{Eps}(q)| \geq MinPts$

Direkt yoğunluk ulaşılabilirlik öz noktaların çiftleri için simetriktir. Ancak genelde, bir öz nokta ve bir sınır noktası söz konusu olduğunda simetrik değildir.

2.2.3.1.3 Yoğunluk ulaşılabilirlik

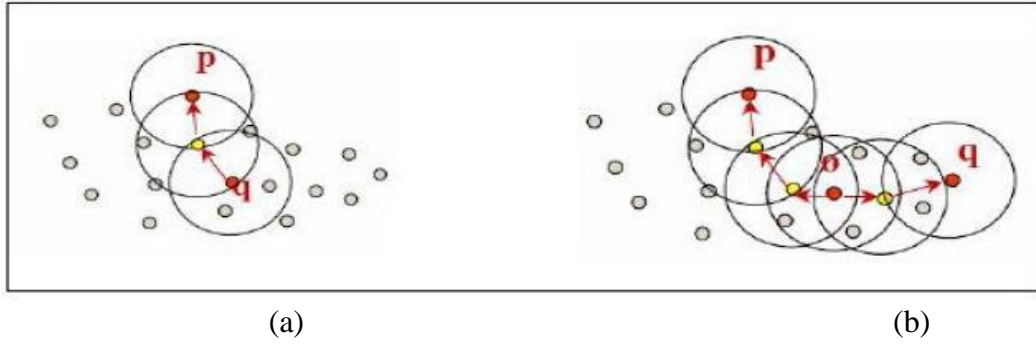
Yoğunluk ulaşılabilirlik, direkt yoğunluk ulaşılabilirliğin kabul edilmiş bir genişlemesidir. Bu ilişki geçişlidir ancak simetrik değildir. Genelde simetrik olmamasına rağmen, yoğunluk ulaşılabilirliğin öz noktalara göre simetrik olduğu açıktır.

Aynı C kümesinin iki sınır noktası birbirinden yoğunluk ulaşılabilir değildir, çünkü öz nokta koşulu her ikisi için de sağlanmayabilir. Ancak C 'de, C 'nin her iki sınır nesnesinin de yoğunluk ulaşılabilir olduğu bir öz nokta olmalıdır. Böylece noktalarının bu ilişkisini içeren yoğunluk-bağlılık kavramını açıklayabiliriz.

2.2.3.1.4 Yoğunluk bağlılık

Eğer Eps ve $MinPts$ koşulları altında o gibi bir noktadan p ve q 'nin ikisi de yoğunluk ulaşılabilir ise p noktası q noktasına yoğunluk bağlıdır.

Yoğunluk bağlılık simetrik bir ilişkidir. Yoğunluk ulaşılabilir noktalar için, yoğunluk bağlılık ilişkisi aynı zamanda dönüşlüdür.



Şekil 2.9 (a) Yoğunluk ulaşılabilirlik ve (b) yoğunluk bağıllık

Bir küme bulmak için DBSCAN keyfi bir p noktası ile başlar ve Eps ile MinPts koşulları altında p 'den yoğunluk ulaşılabilir tüm noktaları elde eder. Eğer p bir öz nokta ise, bu işlem Eps ve MinPts ile bir küme üretir. Eğer p bir sınır nesnesi ise p 'den yoğunluk ulaşılabilir herhangi bir nokta yoktur ve DBSCAN veri tabanının bir sonraki noktasına gider.

Algoritma,

1. Eps ve MinPts parametreleri kullanıcı tarafından verilir.
2. Veri setindeki tüm noktalar sınıflandırılmamış olarak işaretlenir.
3. Eps ve MinPts koşulları altında sınıflandırılmamış bir p öz noktası bulunur. Bu p noktası sınıflandırılmış olarak işaretlenir. Mevcut küme olarak yeni bir küme yaratılır ve p bu mevcut kümeye atanır.
4. p 'nin Eps_Komşuluğundaki tüm sınıflandırılmamış noktalar bulunur. Geçici bir küme yaratılır ve tüm bu noktalar o kümeye konulur.
5. Geçici kümedeki q noktası alınır, sınıflandırılmış olarak işaretlenir, mevcut kümeye atanır ve geçici kümeden kaldırılır.
6. Eps ve MinPts parametreleri ile q 'nun öz nokta olup olmadığına bakılır. Eğer öz nokta ise, q 'nun Eps_Komşuluğundaki sınıflandırılmamış tüm noktalar geçici kümeye eklenir.

7. 5. ve 6. adımlar geçici küme boş olana kadar tekrar edilir.

8. Yeni küme yaratılır ve 3 ile 7 arasındaki adımlar başka öz nokta kalmayınca kadar sürdürülür.

9. Bulunan tüm kümeler işaretlenir, hiçbir kümeye ait olmayan nesnelere sapan olarak atanır.

2.2.3.2 OPTICS

Birçok gerçek veri setinin önemli bir özelliği özel küme yapılarının küresel yoğunluk parametreleri ile karakterize edilememesidir. Çok farklı bölgesel yoğunluklara, veri uzayının farklı bölgelerindeki kümeleri ortaya çıkarmak için ihtiyaç duyulabilir.

Bu gibi bir küme yapısını saptamak ve analiz etmek için ilk alternatif tek bağ metodu gibi bir hiyerarşik kümeleme algoritması kullanmaktır. Ancak bu alternatifin iki dezavantajı vardır. İlki tek bağ etkisidir. Örneğin nesnelere arası küçük bir uzaklık olan birkaç nokta sırası bakımından bağlı kümeler ayrılamazlar. İkincisi hiyerarşik algoritmalar tarafından üretilen sonuçların yani dendogramların anlaşılması veya analizi birkaç yüz nesneden sonra zordur.

İkinci alternatif farklı parametre ayarları ile yoğunluk tabanlı bir ayırıcı algoritma kullanmaktadır. Ancak sonsuz sayıda mümkün parametre değeri vardır. Çok fazla sayıda farklı değerler kullansak bile ki bu her nokta için farklı küme üyeliklerini depolamak için çok fazla ikincil hafıza gerektirir, sonuçların nasıl analiz edileceği açık değildir ve yine ilginç küme seviyelerini kaçırabiliriz.

Bu sorunların çözümlenmesi için, veri setinin her küme seviyesine ilişkin bilgiyi içeren yoğunluk tabanlı küme yapısı hakkında veri tabanının özel bir sıralamasını üreten OPTICS algoritması kullanılmalıdır.

2.2.3.2.1 Yoğunluk tabanlı küme sıralaması

DBSCAN algoritmasını farklı uzaklık parametreleri aynı anda işlem göreceği şekilde genişletebiliriz. Örneğin farklı yoğunluklara ilişkin yoğunluk tabanlı kümeler eş zamanlı

oluşturulabilir. Ancak istikrarlı bir sonuç üretmek için, bir kümeyi genişletirken nesnelerin işlendiği özel bir sıraya uymamız gerekir. Daima en küçük Eps değerine ilişkin yoğunluk ulaşılabilir nesneyi daha büyük yoğunluklu (yani daha küçük Eps’li) kümelerin önce bitirilmesini garantiye almak için seçmeliyiz.

OPTICS algoritması, bir “generating distance” Eps ($0 \leq \varepsilon_i \leq \varepsilon$)’dan daha küçük olan sonsuz sayıdaki ε_i parametreleri için genişletilmiş DBSCAN algoritması gibi bir prensiple çalışır. Burada nesnelerin işleme sırasını ve DBSCAN algoritması tarafından küme üyelikleri atanması için kullanılacak bilgi depolanır. Bu bilgiler her nesne için sadece öz uzaklık ve ulaşılabilirlik mesafesi olmak üzere iki değerden oluşur.

2.2.3.2.2 Bir nesnenin öz uzaklığı

Bir p nesnesinin öz uzaklığı temel olarak p ile onun Eps_Komşuluğu’ndaki bir nesne arasındaki en küçük ε' uzaklığıdır öyle ki p, ε' komşuluğu $N_{\varepsilon}(p)$ 'de yer alıyorsa ε' 'ne göre bir öz nesne olacaktır. Aksi takdirde öz uzaklık tanımlanamaz.

$$\text{Öz_Uzaklık}_{G, \text{MinPts}}(p) = \begin{cases} |N_{\varepsilon}(p)| < \text{MinPts} \text{ ise Tanımsız} \\ \text{aksi takdirde MinPts} - \text{Mesafesi}(p) \end{cases}$$

p: Bir veritabanından nesne

Eps: Uzaklık değeri

$N_{\varepsilon}(p)$: p'nin Eps_Komşuluğu

MinPts: Bir doğal sayı

MinPts – Mesafesi(p): p'den MinPts_Komşusu'na olan uzaklığı

2.2.3.2.3 Ulaşılabilirlik mesafesi

Bir p nesnesinin bir başka o nesnesine göre ulaşılabilirlik mesafesi en küçük uzaklıktır, öyle ki o bir öz nesne ise p o'dan direkt yoğunluk ulaşılabilir. Bu durumda ulaşılabilirlik mesafesi o'nun öz uzaklığından daha küçük olamaz çünkü daha küçük değerler için hiçbir nesne hiçbir nesne o'dan direkt yoğunluk ulaşılabilir olamaz. Aksi takdirde, o bir öz nesne değil ise Eps'de olsa bile o'ya göre p'nin ulaşılabilirlik mesafesi tanımlanamaz. Bir p nesnesinin ulaşılabilirlik mesafesi hesaplanacağı öz nesneye bağlıdır.

OPTICS algoritması veritabanının bir sıralamasını oluşturur, ek olarak öz uzaklığı ve uygun ulaşılabilirlik mesafesini her nesne için depolar. Bu bilgi Eps'den küçük herhangi bir ε mesafesi ile tüm yoğunluk tabanlı kümeleri genişletmek için yeterlidir.

OPTICS Algoritması:

OPTICS algoritması iki bölümden oluşur;

1. Bölüm: Sıralanmış küme yapısının oluşturulması,

1. ε_0 mesafesi ile MinPts kullanıcı tarafından verilir.

2. Boyutu veri setindeki noktaların sayısı kadar olan bir vektör oluşturulur. Bu vektörde, veri setindeki her nokta kendisine uygun bir elemana sahip olacaktır. Her eleman noktanın sıra numarasını, p_1 'in öz mesafesini ve p_1 'den en yakın öz noktaya olan ulaşılabilirlik mesafesini vektörde tutar. Bu vektör ile sıralanmış küme yapısı elde edilecektir. Başlangıçta boştur.

3. Bilgileri sıralanmış yapıda tutulmayan isteğe bağlı bir nokta seçilir ve ε_0 ile MinPts parametrelerine göre öz nesne olup olmadığına bakılır. Eğer bir öz nesne değilse, öz mesafesi ve ulaşılabilirlik mesafesi tanımlanamaz olarak atanır. Eğer bir öz nesne ise, öz mesafesi hesaplanır ve ulaşılabilirlik mesafesi mevcut seçilen nokta kendisinden önce sıralanmış yapıda tutulan herhangi bir noktadan direkt yoğunluk ulaşılabilir olmadığı için tanımlanamaz olarak atanır. Şimdi mevcut seçilen noktanın bilgileri sıralanmış yapıda saklanır.

4. 3. adım p öz nesnesi bulununcaya kadar tekrar edilir.

5. ε_0 ve MinPts parametrelerine göre p'nin ε komşuluğundaki tüm noktalar bulunur. Kaynakların sıralanmış bir kümesi yaratılır ve bu ε komşuluğundaki tüm noktalar kümeye konulur. Kümedeki tüm noktalar p'ye olan ulaşılabilirlik mesafelerine göre küçükten büyüğe doğru sıralanırlar.

6. Kümedeki ilk kaynak q alınır, sıralanmış yapı için q 'nun öz mesafesi ve p 'ye olan ulaşılabilirlik mesafesi dahil tüm bilgileri eklenir. Ardından kaynaklar kümesinden q kaldırılır. Eğer q bir öz nokta ise, sıralanmış yapıda gözükmeyen q 'nun tüm ε komşuları kaynaklar kümesine eklenir. Eğer herhangi bir r noktası zaten kümedeyse, r 'den p 'ye olan ulaşılabilirlik mesafesi ile r 'den q 'ya mesafe karşılaştırılır. Sonuçta daha küçük olan değer alınır. Mevcut ulaşılabilirlik mesafelerine göre kaynaklar tekrar sıralanır.

7. Kaynaklar kümesi boşalana kadar 6. adım tekrar edilir.

8. Veri setindeki tüm noktaların bilgileri sıralanmış yapıya kaydedilene kadar adım 3'den adım 7'ye kadar tekrar edilir.

2. Bölüm: Sıralanmış küme yapısından kümelerin çıkarılması,

1. ε , kullanıcı tarafından verilir.

2. Küme numarası 1 olarak ayarlanır.

3. Sıralanmış yapıdan ε ve MinPts parametreleri ilk öz nokta p alınır. Küme numarası ile tanımlanmış mevcut kümeye atanır.

4. Yapıda p 'nin arkasından gelen tüm noktalar, eğer ulaşılabilirlik mesafeleri ε 'dan daha büyük değil ise mevcut kümeye atanırlar. Bu adım yapıdaki ulaşılabilirlik mesafesi ε 'dan daha büyük bir q noktası bulunduğu anda sona erer.

5. Küme numarası 1 arttırılır.

6. q noktasından başlayarak yapıyı taramaya devam edilir, ikinci kümeyi elde etmek için 3-5 adımları uygulanır.

7. Yapıdaki tüm noktalar işleninceye kadar algoritma yeni kümeler bulmaya devam eder.

8. Kümeler bulunmuştur. Eğer bir nokta herhangi bir kümeye ait değil ise sapandır.

2.2.3.3 FJP metodu (bulanık düğüm noktaları)

FJP metodunda, başlangıç kümesinin her elemanı çok boyutlu uzayın bir bulanık bir noktası olarak ele alınır. Bu metodun diğer bulanık kümeleme metotlarından temel farkı, bulanıklığı hiyerarşik bakış açısıyla değerlendirmesidir. Kümelemenin bulanıklığı, göz önünde bulundurulmuş özelliklerin ne kadar detaylı ele alındığı ile ifade edilir. Minimum bulanıklık derecesi sıfırda, tüm nesnelere birbirlerine benzemez olacakları için, her nesne ayrı bir küme olarak düşünülür. Maksimum bulanıklık derecesinde ise, tüm elemanlar bir kümeye ait olacaklardır. Elemanlar $[0,1]$ aralığındaki üyelik derecelerini alarak birbirlerine benzerliklerine göre kümelere atanırlar (Nasibov 2005).

2.2.4 Grid (ızgara) bazlı kümeleme metodu

Grid bazlı metot, kümeleme yapılacak alanın sonlu sayıda hücelere bölünmesiyle oluşur. Bu metodun ana avantajı, genelde birbirinden bağımsız sayıda veri nesnelere hızlı işlem zamanıdır.

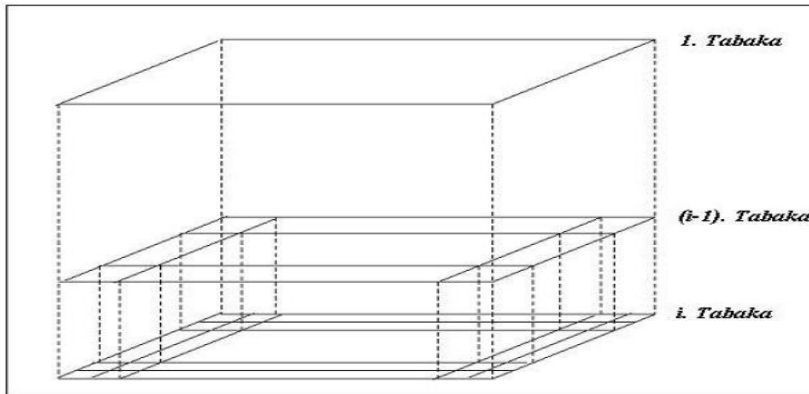
2.2.4.1 STING (istatistiksel bilgi gridi)

STING, grid hücrelerindeki istatistiksel bilgiyi araştırır. STING, hiyerarşik yapıyı kullanarak uzayı dikdörtgen hücelere bölen bir tekniktir. STING algoritması, hücrelerin içindeki bireylerin sayısal özelliklerinin her birinin ortalama, standart sapma, varyans, minimum, maksimum gibi istatistiksel parametrelerini hesaplar. Sonra farklı seviyelerdeki kümeleme bilgisini temsil ettiği gibi, ızgara hücrelerinin hiyerarşik yapısını oluşturur. STING, yeni bir bireyin kümelere verimli tahsisini veya sorgulama için kümeleme bilgisinin kullanımını kolaylaştırır.

STING'in bazı avantajları;

- Veriler hakkındaki özet bilgiyi temsil eden her bir hücrede depolanan istatistiksel bilgidir, hesaplanması sorgulamadan bağımsızdır.
- Grid yapısı, paralel işleme ve güncelleştirmelere uygundur.

- Metodun verimi asıl avantajıdır. STING, hücrelerdeki istatistiksel parametreleri hesaplamak için veri tabanına bir kere gider, kümeleri oluşturma zaman karmaşıklığı $O(n)$ 'dir, n nesnelerin toplam sayısıdır. Hiyerarşik yapıyı oluşturduktan sonra, sorgu işleme zamanı $O(g)$, g en alt seviyedeki hücre sayısıdır (genellikle $n > g$ olur).



Şekil 2.10 STING kümeleme için hiyerarşik yapı

STING'in kalitesi grid yapısındaki en alt seviye taneciğe bağlıdır. Tanecikler hassas ise işlem maliyeti artar, bununla beraber, en alt seviyedeki grid yapısının kalınlığı kümeleme analiz kalitesini azaltabilir. STING, ana hücrenin çocuk ve komşuları ile ilişkilerini göz önünde bulundurmaz. Kümeleme sınırları düşey veya yataydır, diyagonal sınır yoktur. Bu kalite ve doğruluğu azaltır.

2.2.4.2 WaveCluster (wavelet dönüşümü kullanarak kümeleme)

Wavecluster, uzaysal veriyi frekans alanına dönüştürmek için ve dönüştürülen uzaydaki yoğun olan bölgeleri bulmak için işaret işleme tekniklerini kullanır. İlk olarak veri uzayını çok boyutlu grid yapısına dönüştürür. Sonra wavelet dönüşümü kullanarak yoğun bölgeleri bulur ve uzayda dönüşüm yapar.

Wavelet model, n defa dönüşüm yaparak bir boyutlu sinyali n boyutlu işaretlere dönüştürebilir. Farklı çözüm seviyelerindeki göreceli mesafe verileri wavelet dönüşümü sayesinde saklanır. Bu doğal kümeleri daha çok ayırt edilebilir hale getirir. İlgi alanındaki yoğun bölgeler arayarak kümeler tanımlanabilir (Sheikholeslami vd.1998).

Avantajları;

- Bu metot danışmansız kümeleme sağlar ve noktaların yoğun olduğu bölgeleri ortaya çıkarmak için çizgi – biçimli (line – shaped) filtre kullanır. Bu özelliği sayesinde aynı zamanda küme sınırları aynı zamanda dışındaki zayıf bilgilerin ortadan kalkmasını sağlar.
- Çoklu – ayırıştırma (multi – resolution) özelliği, farklı doğruluk seviyelerindeki kümeleri ortaya çıkarmada yardımcı olur.
- n veri tabanındaki bireylerin sayısı olmak üzere, zaman karmaşıklığı $O(n)$ 'dir.
- Keyfi şekilli kümeleri ortaya çıkarmada ve büyük veri setlerini işlemede başarılıdır.
- Aykırı gözlemler olması durumunda da başarılı sonuçlar verir.
- Küme sayısı ya da komşuluk yarıçapı gibi girdi parametrelerine gerek duymaz.

Wavelet tabanlı kümeleme çok hızlıdır. Algoritmanın paralel uygulaması mümkündür.

2.2.4.3 CLIQUE (yüksek boyutlu alanda kümeleme)

CLIQUE, yoğunluk tabanlı ve grid tabanlı kümelerin birleşmesinden oluşan bir algoritmadır. Büyük veritabanlarında yüksek boyutlu veri kümelemek için yararlıdır.

Verilen büyük kümeli çok boyutlu veri noktalarında, veri alanı genelde tek biçimli olmaz. CLIQUE kümeleme, dağınık veri kümelerinde seyrek ve kalabalık bölgeleri tanımlar. Toplam veri noktalarının parçası giriş model parametresini aşarsa bölüm yoğundur. CLIQUE'da, küme bağlantılı yoğun bölümün maksimal bir kümesi olarak tanımlanır.

CLIQUE iki adımda çalışır;

İlk adımda CLIQUE, üst üste binmeyecek şekilde n -boyutlu dikdörtgen şeklinde parçalara ayırır. Alt alanlar, yoğun bölümlerin kesişmesi durumunda bizim arama yapmak için aday olması anlamına gelir.

Aday arama alanının tanımlanması Apriori özelliğine dayanır. Genelde özellik, arama alanındaki öncelikli bilgiyi kullanır, böylece alan budanarak bölünür.

CLIQUE için uyarlanan özellik, şu şekilde açıklanır: Eğer k -boyutlu bölüm yoğun ise $k-1$ boyutlu bölüme bakılır. $k-1$ boyutlu bölümde yoğunluk yoksa k -boyutlu bölümde de yoğunluk olmaz. Bu nedenle, $k-1$ boyutlu alanların yoğun bölümlerinde bulunan, k boyutlu alanların yoğun bölümleri içinde potansiyel veya aday yoğunlukları üretebiliriz. Genelde sonuçlanan alan, orijinal alandan daha küçük olacaktır.

İkinci adımda, CLIQUE her küme için minimal bir tanımlama yapar. Her küme için maksimum bölge tanımlanır. Maksimum bölge bağlantılı yoğun bölümlerin kümesini kapsar. Sonra her küme için minimal kapsamı tanımlar.

CLIQUE otomatik olarak en yüksek boyutlu alt alanları bulduğu gibi her alt alandaki yüksek yoğunlukta kümeleri de bulur. Girilen satırların sırasına duyarlıdır ve herhangi bir geleneksel veri dağılımını tahmin edemez. Giriş boyutuna göre lineer ölçeklenir, verideki boyut sayısı arttıkça iyi ölçeklenme özelliğine sahiptir. Basitlik masrafına göre kümeleme doğruluğu düşebilir.

3. ÖNCEKİ ÇALIŞMALAR

3.1 Genel

Kümeleme analizi, eldeki veri setinin çeşitli parametrelere göre kümelere ayrılarak incelenmesine olanak sağlamaktadır. Bu, eldeki veri setinin daha kolay anlaşılması ve daha iyi yorumlanması açısından büyük bir avantaj sağlamaktadır. Bu yüzden kümeleme analizi birçok çalışma alanında kullanılmaya başlanmıştır. Bu çalışma alanlarından bazıları;

İş: İş sektöründe, kümeleme, müşteri veritabanlarına göre önemli pazar gruplarını belirlemeye yardımcı olabilir.

Biyoloji: Biyolojide, benzer fonksiyonel özellikleri taşıyan genleri kategorize etmede ve popülasyonun doğasını kavramada kullanılabilir.

Uzayla İlgili Veri Analizi: Çeşitli yöntemlerle elde edilebilen çok büyük miktarlardaki uzaysal verinin anlaşılmasını ve analizinin yapılmasını otomatik hale getirebilir.

Web Madenciliği: Burada kümeleme web üzerindeki belgelerin anlamlı gruplarını keşfetmede kullanılabilir.

Sosyal Network Analizi: Sosyal network çalışmasında, büyük insan gruplarındaki toplulukları ayırt etmede kullanılabilir.

Görüntü Segmantasyonu: Kümeleme, dijital bir görüntüyü sınır belirleme veya nesne tanıma için farklı bölgelere bölmek için kullanılabilir.

Kayıc Harita Optimizasyonu: Flickr veya Google haritaları ve diğer harita siteleri kümelemeyi, harita üzerindeki işaret sayılarını azaltmak için kullanır. Bu siteyi daha hızlı yapar ve görsel dağınıklığı azaltır.

3.2 Daha Önce Yapılmış Çalışmalar

Karpat ve Yılmaz (1997), Türkiye’deki trafik kaza verilerini incelemişler ve trafik kaza verilerini iller bazında oluş şekillerine göre kümelendirmişlerdir. Bu çalışma sonucu ortaya çıkan çarpışma ve yayaya çarpma bakımından illerin kümelenmesi Tablo 3.1’de verilmiştir.

Tablo 3.1 Çalışma sonucu elde edilen çarpışma ve yayaya çarpma bakımından illerin kümelenmesi (Karpat ve Yılmaz 1997)

Kümelere	İller
1.	Adana
2.	Adıyaman, Afyon, Ağrı, Amasya, Artvin, Bilecek, Bingöl, Bitlik, Bolu, Burdur, Çanakkale, Çankırı, Çorum, Diyarbakır, Edirne, Elazığ, Erzincan, Erzurum, Giresun, Gümüşhane, Hakkari, Isparta, Kars, Kastamonu, Kırklareli, Kırşehir, Malatya, K.Maraş, Mardin, Muş, Nevşehir, Niğde, Ordu, Rize, Siirt, Sinop, Sivas, Tekirdağ, Tokat, Trabzon, Tunceli, Ş.Urfa, Uşak, Van, Yozgat, Aksaray, Bayburt, Karaman, Kırıkkale, Batman, Şırnak, Bartın, Ardahan, Iğdır, Yalova, Karabük, Kilis, Osmaniye
3.	Ankara
4.	Antalya, Bursa, Konya
5.	Aydın, Balıkesir, Denizli, Eskişehir, G.Antep, Hatay, İçel, Kayseri, Kocaeli, Kütahya, Manisa, Muğla, Sakarya, Samsun, Zonguldak
6.	İstanbul
7.	İzmir

Tablo 3.1 incelendiğinde Adana, Ankara, İstanbul ve İzmir’in taşıt sayısı, kaza sayısı, ölü sayısı ve yaralı sayısı bakımından bir ilden oluşan farklı kümelere yer aldığı, ayrıca Antalya, Bursa ve Konya’nın ortak bir kümede toplandığı açıkça görülmüştür (Karpat ve Yılmaz 1997).

Karavaşin ve Saplıoğlu (2005), coğrafi bilgi sistemi yardımı ile Isparta kenti, kentiçi trafik kazalarını analiz etmişlerdir. Çalışma sonucunda, son beş yılda Isparta’ da kavşaklar ile cadde ve sokaklarda oluşmuş kaza sayıları tespit edilmiş, kavşaklarda beş yılda kayıtlara geçen 1016 kazanın kara noktaları belirlenmiştir.

Geurts ve Wets (2003), çalışmalarında kara nokta ve kara bölgelerin analizinde kullanılan metot ve teknikleri literatür olarak gözden geçirmişler, kara noktaları tanımlama ve sıraya koymada kullanılan birkaç alternatif metot üzerinde durmuşlardır.

Sonuç olarak Belçika, Danimarka ve Avustralya' da kara noktaları analiz etmede kullanılan bu metot ve teknikler hakkında genel bir bakış sunmuşlardır.

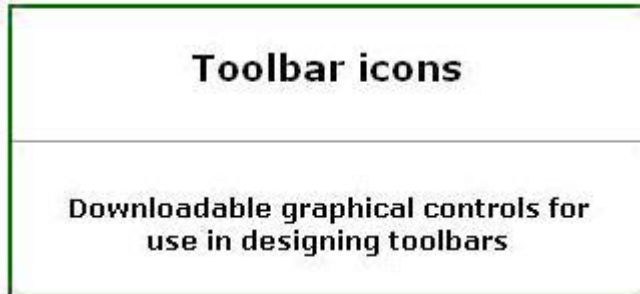
Kara nokta iyileştirme programı, kazaların meydana gelişlerini duyarlı olarak ele alabilen bir güvenlik geliştirme programı tipidir. Aslında, bu tarz programların uygulaması, gelişen ülkelerde mevcut olmayan veya limitli olan, ilgili kaza verilerine ihtiyaç duymaktadır. Bunun için Kowtanapanich vd. (2005), çalışmalarında bu engelin üstesinden gelen destekleyici bir yaklaşımı (halk katılımı yaklaşımı) ele almışlardır. Amaçlanan, Kaza Halk Katılım Programı çerçevesi sayesinde, kara noktaların tanımlanmasına katkıda bulunmada bir halk katılımı yaklaşımının nasıl kullanılabileceğini ispatlamaktır.

Grip A virüsünün hemaglutinin geninin sürekli mutasyonları, yıllık salgınlara sebep olan yeni antijenik türleri meydana getirir. Dushoff vd. (2003), son yirmi yılda kümeleme analizini, hemglutinin gelişiminin yapısı ve hareketi üzerinde kullanarak, hemaglutinin serilerinin kümeler içine toplandığı, amino-asid boşluğunda, kritik bir uzunluk bulmuşlardır ve bu kümelerin spatio-temporal dağılımını incelemişlerdir.

Bir web sitesinin organizasyonel yapısı, o sitenin kolay kullanımının başlıca unsurudur. Ziyaretçiler, site tarafından rahat bir şekilde ve güvenli olarak yönlendirilebilmelidirler. Maalesef çok sayıda tüzel web sitesi kendi yapılarını, şirketin çeşitli bölümlerinin dahili düzeninden miras almaktadır. Bu tarz tasarımlar site kullanıcılarına çoğu kez anlaşılmaz gelmektedir.

Martin (2001), kümeleme analizini, card-sorting (kullanıcıların maddeler arasındaki ilişki algılamasını anlayabilmek için yararlı olan bir veri toplama metodu) verisini kümelemek amacıyla kullanmıştır. Bu, web site sayfalarının organizasyonunda, çoklu katılımcı girdisine anlam yükleyebilmek için uygulanan elverişli bir metottur. Burada kümeleme analizi, her bir olası kart çifti üyelerinin ortak bir kümede ne sıklıkla görüldüğüne dayanan kart çiftleri arasındaki algılanan ilişkilerin kuvvetini hesaplama yöntemiyle card-sorting (kart-sınıflama) verisinin miktarını belirler.

Özetle bu çalışmada, kart-sınıflama uygulamaları ve kümeleme analizinin tasarımcılara, hedef kullanıcı kitlesinin içerik organizasyonu için beklentilerini anlamada yardımcı olduğu açıkça anlaşılmıştır.



Şekil 3.1 Kart-sınıflama görevi için örnek sayfa kartı

Huang vd. (2004), RNA döngü yapılarının kıyaslanması ve sınıflandırılması için bir hesap yaklaşımı geliştirmişlerdir. Burada, atomik kararlı RNA yapılarında tespit edilen firkete veya içten döngüleri uygunluk karşılaştırması ile kıyaslamışlar ve RNA'nın ilgili kısımlarındaki bütün çiftler arasındaki kök-ortalama-kare sapma (KOKS) değerlerini, farklı moleküllerden olsa bile hesaplayarak, sonuç matrisi olan KOKS uzaklıkları üzerinde kümeleme analizini uygulamışlardır. Sonuç olarak bu çalışmada kümeleme analizi, objektif olarak birbirine benzeyen kıvrım gruplarını ortaya çıkarmıştır.

Espona vd. (1995), çalışmalarında kümeleme analizi tekniklerini Sınıf I maloklüzyonun (kapanış bozukluğu) istatistiksel alttürlerini elde etmek, farklı kümelerden kliniksel ve sefalometrik karakteristikleri belirlemek, cinsiyetin ve yaşın gruplama örnekleri üzerindeki etkisini analiz etmek için kullanmışlardır. Burada kümeleme analizi, her hastadan alınan, her iki kliniksel ve sefalometrik 20 veri aracılığıyla sağlanan bilgiye uygulanmıştır. Sonuç olarak ise, sınıf I maloklüzyon' daki gruplama örneğinin, daha çok genç yaş düzeyinde tanımlanmış formlarda görüldüğü ve bunun yaş ilerledikçe kaybolduğu saptanmıştır. Ayrıca kümeleme, Sınıf I maloklüzyonun erkek ve kadınında çok benzer olduğu gözlemlenmiştir.

Cebeci (2004), Türkiye' deki üniversitelerin web tabanlı yaptıkları yayınlara göre derecelendirilmesi çalışmasında kümeleme analizini kullanmış ve üniversiteleri, toplam yayın sayıları bakımından beş gruba ayırmıştır. Çalışma sonucunda

Hacettepe Üniversitesi' nin açık şekilde önde olup ilk grup içinde yer alan tek üniversite olduğu ve ikinci grubu Ankara ve İstanbul Üniversiteleri' nin oluşturduğu görülmüştür.

Tablo 3.2 Çalışmada elde edilen kümeleme analizi sonuçları (Cebeci 2004)

Ölçüt	Grup 1	Grup 2	Grup 3	Grup 4	Grup 5
WoS	Hacettepe	Ankara, İstanbul	Gazi, Ege, İTÜ, ODTÜ	Atatürk, Dokuz Eylül, Bilkent, Marmara, Boğaziçi, Çukurova, Uludağ, İnönü, KTÜ, Fırat, Akdeniz, Başkent, Selçuk, Ondokuz Mayıs, Erciyes	Diğer Üniversiteler
GS	Hacettepe	Ankara, ODTÜ, İstanbul	Gazi, Bilkent, İTÜ, Ege	Atatürk, Çukurova, Erciyes, Marmara, Selçuk, Dokuz Eylül, Boğaziçi	Diğer Üniversiteler
WBYG	ODTÜ, Bilkent	Ankara, İTÜ, Boğaziçi, Uludağ, Anadolu, Çukurova, İstanbul Bilgi, Ege, Dokuz Eylül, Gazi	Diğer Üniversiteler		

Avrupa Birliği günümüzün en güçlü yapılarından biridir. Halen içinde Türkiye'nin de bulunduğu birkaç Avrupa ülkesi Avrupa Birliği' ne üye olabilmek için gayret göstermekte, bu doğrultuda birçok reformlar yaparak Avrupa Birliği standartlarını yakalamaya çalışmaktadır. Turanlı vd. (2006), çalışmalarında Avrupa Birliği üye ülkeleri ile aday üye ülkeler arasında var olan ekonomik benzerlikleri ortaya koyarak, aday ülkelerin üye olmak için yeterli olup olmadıklarını anlamaya çalışmışlardır. Amacın, Avrupa Birliği aday ve üye ülkelerinin ekonomik benzerliklerini ortaya koyulması olmasından dolayı Kümeleme Analizinin kullanılmasını uygun bulmuşlardır. Elde edilen sonuç Tablo 3.3' de gösterilmiştir.

Çalışmanın sonucunda, Türkiye'nin ekonomik olarak Avrupa Birliği'ne üye olmaması için bir neden görülmediğini, üyelikte yaşanan sorunların siyasi ve politik sorunlar olduğunu tespit etmişlerdir.

Tablo 3.3 Çalışma sonucu elde edilen kümeleme tablosu (Turanlı vd. 2006)

Küme Sayısı	Küme Numarası	Küme Elemanları
K= 2	Birinci Küme	Belçika, Danimarka, Almanya, İspanya, Fransa, İrlanda, İtalya, Lüksemburg, Hollanda, İsveç, Avusturya, Finlandiya, İngiltere
	İkinci Küme	Çek Cumhuriyeti, Estonya, Yunanistan, Kıbrıs Rum Kesimi, Letonya, Litvanya, Polonya, Macaristan, Malta, Portekiz, Slovenya, Slovakya, Bulgaristan, Hırvatistan, Romanya, Türkiye
K= 3	Birinci Küme	Belçika, Danimarka, Almanya, İspanya, Fransa, İrlanda, İtalya, Malta, Hollanda, İsveç, Avusturya, Slovenya, Finlandiya, İngiltere
	İkinci Küme	Çek Cumhuriyeti, Estonya, Yunanistan, Kıbrıs Rum Kesimi, Letonya, Litvanya, Polonya, Macaristan, Portekiz, Slovakya, Bulgaristan, Hırvatistan, Romanya, Türkiye
	Üçüncü Küme	Lüksemburg
K= 4	Birinci Küme	Lüksemburg
	İkinci Küme	Çek Cumhuriyeti, Yunanistan, Kıbrıs, Rum Kesimi, Malta, Portekiz, Slovenya, İspanya
	Üçüncü Küme	Belçika, Danimarka, Almanya, Fransa, İrlanda, İtalya, Hollanda, İsveç, Avusturya, Finlandiya, İngiltere
	Dördüncü küme	Estonya, Letonya, Litvanya, Polonya, Macaristan, Slovakya, Bulgaristan, Hırvatistan, Romanya, Türkiye

4. ÇALIŞMA ALANI VE VERİ

4.1 Genel

Bu bölümde, giriş bölümünde de bahsettiğimiz üzere Denizli kentine ait 2004, 2005 ve 2006 yıllarına ait trafik kaza verileri, NCSS 2007 ve Matlab programları yardımı ile kümeleme analizi yöntemlerinden olan k-ortalamalar ve bulanık kümeleme metotları kullanılarak kümelere ayrılmıştır. Burada koordinat verileri her iki k-ortalamalar ve bulanık kümeleme analizine, kazaya etki eden diğer faktörlerin verileri ise sadece k-ortalamalar kümeleme analizine tabi tutulmuştur. Analizin daha verimli olması açısından veriler MS Excel ortamında incelenmiş, eksik ya da kaydedilmemiş veriler ile yanlış kaydedilmiş veriler elemeyen geçirilmiştir.

4.2 Çalışma Alanı ve Kullanılabilir Veri

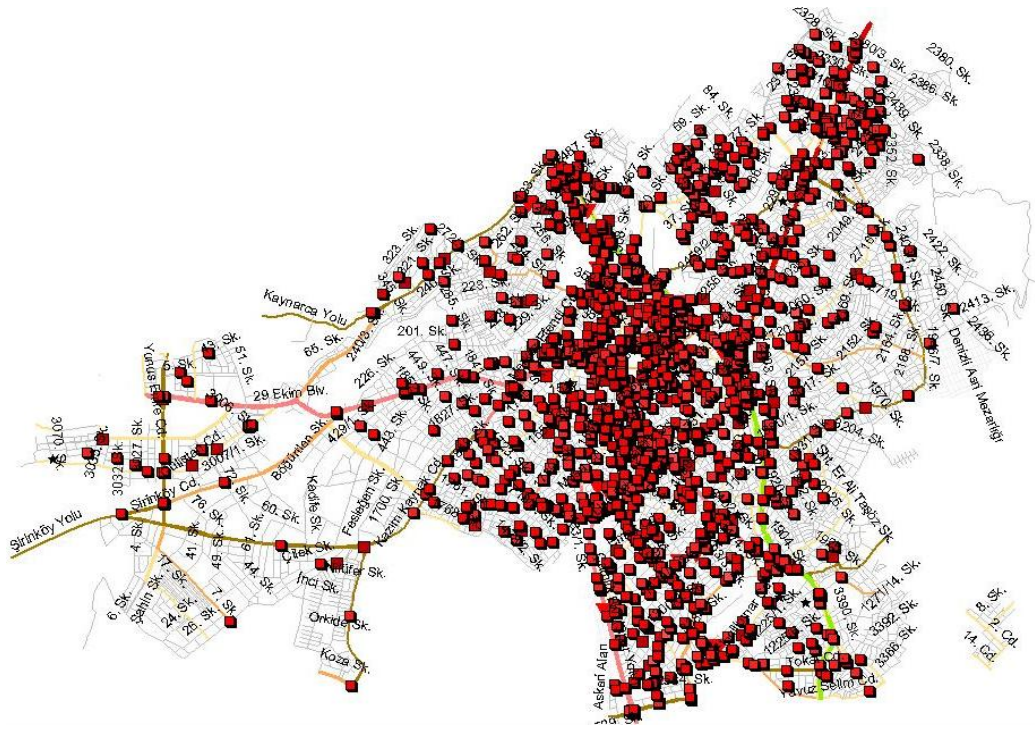
Bu çalışmada, Denizli yerel polisi tarafından kayıt edilen, kentin 2004, 2005 ve 2006 yıllarına ait trafik kaza verileri kullanılmıştır. Trafik kaza raporunda yer alan bilgiler ise şu şekildedir;

- Kazanın yeri ve zamanı
 - Zaman ve tarih
 - Yol Türü
 - Yol Bölünme
 - Yerleşim Durumu
- Kaza Türü
 - Araç sayısına göre kaza türü
 - Oluşumuna göre kaza türü
- Hava ve Gün Durumu
 - Hava durumu
 - Gün durumu

- Yol ve çevre özellikleri
 - Trafik lambası
 - Aydınlatma
 - Yol şerit çizgisi
 - Yaya kaldırımı
 - Banket
 - Trafik işaret levhası
 - Yolda Çalışma
 - Trafik Görevlisi
 - Görüşe engel cisim
 - Kaza sonrası araç haricinde hasar gören diğer unsurlar
 - Yolda yön
 - Yolun kaplama cinsi
 - Yolun yüzeyi
 - Yoldan kaynaklanan sorunlar
 - Yol sorununa ait uyarıcı işaretleme
 - Yolun geometrik özelliği
 - Yatay güzergah
 - Düşey güzergah
 - Kavşak
 - Geçitler
 - Diğerleri
- Kazaya karışan araçlar
- Kazaya karışan sürücüler
- Konum bilgisi

İlk olarak, bütün bu veriler MS Excel programı kullanılarak bilgisayar ortamına aktarılmıştır. Daha sonra, MAPINFO yazılımında bulunan sokak tanımlama sistemi kullanılarak her bir kaza noktasının koordinatı belirlenmiştir. Son olarak kaza

noktalarının koordinatlarını da içeren bu veriler, Excel veri tabanından MAPINFO veri tabanına transfer edilmiştir.



Şekil 4.1 Denizli kenti 2004, 2005 ve 2006 yıllarına ait trafik kaza verisinin harita üzerindeki dağılımı

4.3 NCSS Hakkında

NCSS (Number Cruncher Statistical System) gelişmiş, kolay kullanımı olan çok seçenekli bir analiz programıdır.

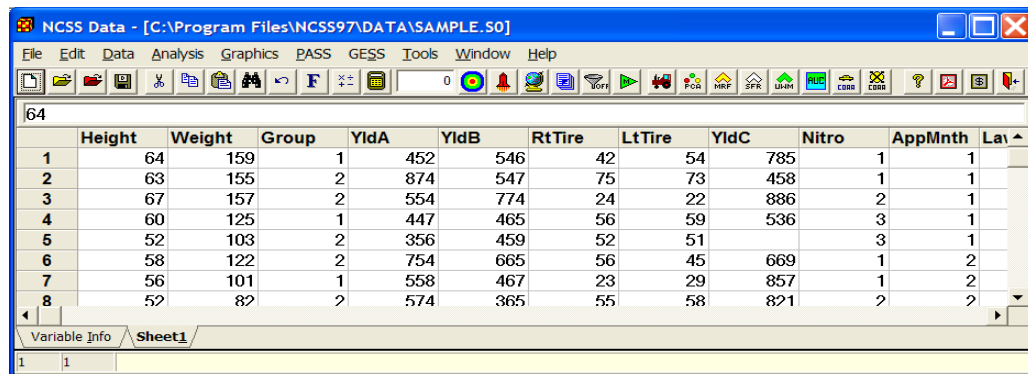
NCSS üç ana pencere tarafından kontrol edilir:

- Veri Penceresi
- İşlem Penceresi
- Çıktı Penceresi

Her bir menünün kendine ait menü çubuğu ve araç çubuğu bulunmaktadır. Şimdi üç ana pencerenin her birini kısaca tanımlayacağız.

4.3.1 NCSS veri penceresi

NCSS veri penceresi o anda analiz edilen veriyi içerir. Bu pencere verinizi görüntülemenize, değiştirmenize ve kaydetmenize olanak verir. Hesap tablosu görünümü ve hissi uyandırır. Bu ana NCSS penceresidir. Bu pencereyi kapatmak NCSS programını sonlandıracaktır.

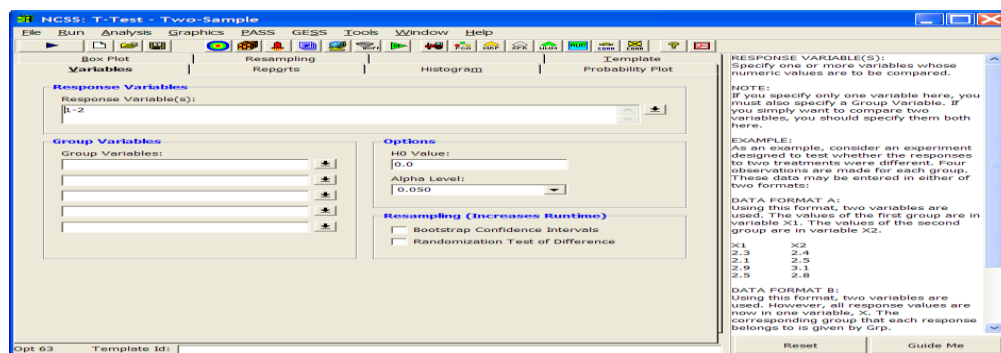


	Height	Weight	Group	YldA	YldB	RtTire	LtTire	YldC	Nitro	AppMnth	Lax
1	64	159	1	452	546	42	54	785	1	1	
2	63	155	2	874	547	75	73	458	1	1	
3	67	157	2	554	774	24	22	886	2	1	
4	60	125	1	447	465	56	59	536	3	1	
5	52	103	2	356	459	52	51		3	1	
6	58	122	2	754	665	56	45	669	1	2	
7	56	101	1	558	467	23	29	857	1	2	
8	52	82	2	574	365	55	58	821	2	2	

Şekil 4.2 NCSS veri penceresi

4.3.2 NCSS işlem penceresi

NCSS işlem penceresi detaylı bir analiz için opsiyonları ayarlamanıza olanak verir. Bir çoklu regresyon, bir ANOVA veya bir dağılım grafiği çalıştırıyorsanız da, bu işlemin opsiyonunu işlem penceresinden ayarlayacaksınız. Bu pencereyi kapatmanız NCSS programını sonlandırmayacaktır.



RESPONSE VARIABLE(S): Specify one or more variables whose numeric values are to be compared.

NOTE: If you specify only one variable here, you must also specify a Group Variable. If you simply want to compare two variables, you should specify them both here.

EXAMPLE: As an example, consider an experiment designed to test whether the responses to two treatments were different. Four observations are made for each group. These data may be entered in either of two formats:

DATA FORMAT A: Using this format, two variables are used. The values of the first group are in variable X1. The values of the second group are in variable X2.

X1	X2
2.3	2.4
2.1	2.5
2.9	3.1
2.5	2.8

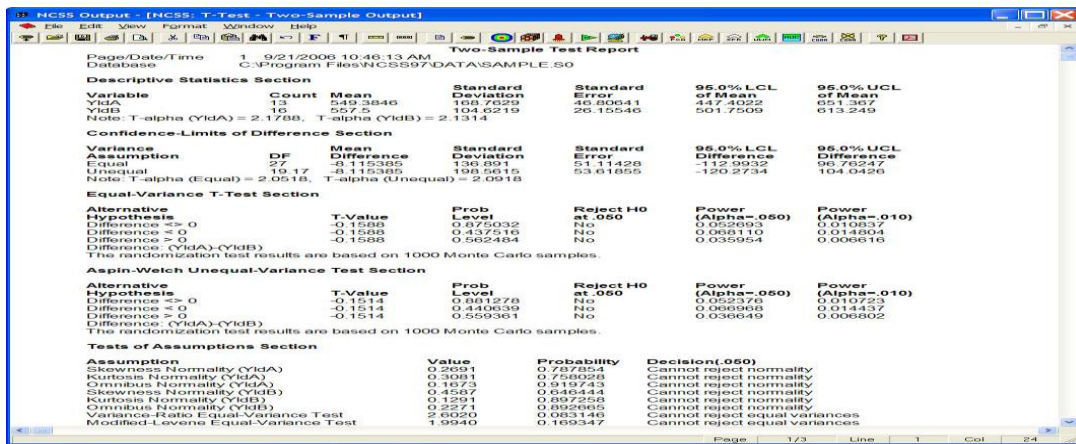
DATA FORMAT B: Using this format, two variables are used. However, all response values are now in one variable, X. The corresponding group that each response belongs to is given by Grp.

Şekil 4.3 NCSS işlem penceresi

4.3.3 NCSS çıktı penceresi

NCSS çıktı penceresi istatistiksel ve grafik işlemlerinden olan çıktıları gösterir. Küçük bir word-işlemcisi gibi çalışarak çıktılarınızı görüntülemenize, yönetmenize,

kaydetmenize ve yazdırmanıza olanak tanır. Bu pencereyi kapatmanız NCSS programını sonlandırmayacaktır.



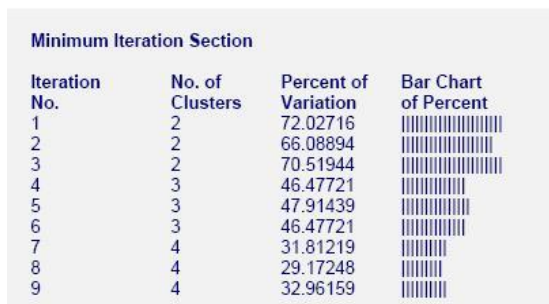
Şekil 4.4 NCSS çıktı penceresi

4.3.4 NCSS ile k-ortalamalar kümeleme analizi

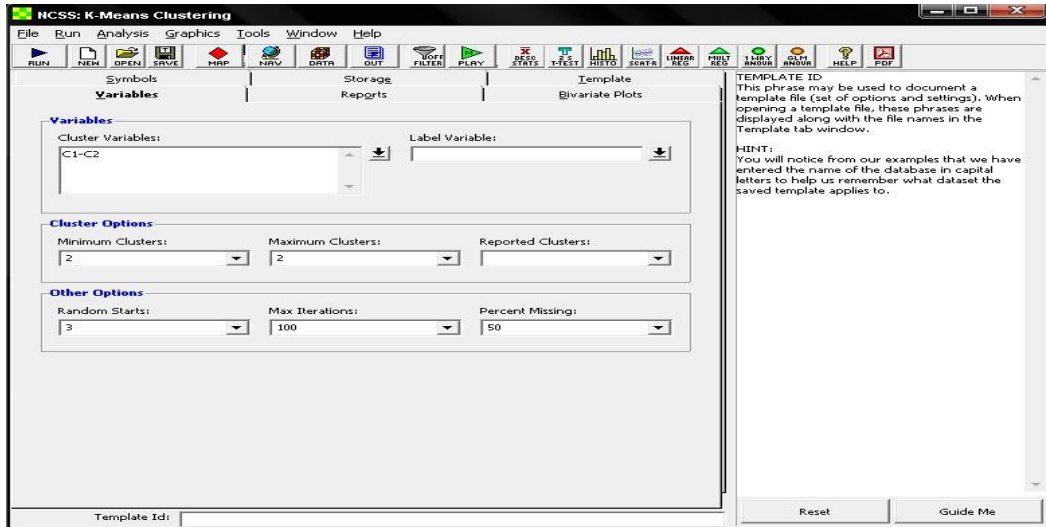
K-ortalamalar algoritması oldukça büyük veri setlerine uygulanabildiği için popüler bir algoritmadır. K-ortalamalar algoritması bütün verilerin sürekli olduğunu varsaymaktadır.

İlk etapta optimum küme sayısı bilinmediği için, veri setinin iki kere analize tabi tutulması gerekir. İlk analizde, maksimum küme sayısına makul seviyede olacak şekilde rasgele bir küme sayısı girilir ve elde edilen analiz çıktısı incelenir.

Elde edilen çıktıda ilk bakılması gereken yer minimum iterasyon kısmında (minimum iteration section) yer alan değişim yüzdesidir (percent of variation). Burada değişimin çok açık bir şekilde değişmediği küme sayısı optimum küme sayısı olarak seçilir.



Şekil 4.5 NCSS minimum iterasyon çıktısı



Şekil 4.6 NCSS k-ortalamlar kümeleme analizi işlem penceresi

İkinci analizde maksimum küme sayısı ve raporlanacak küme (reported cluster) sayısına bu elde edilen optimum küme sayısı girilerek analiz sonuçlandırılır.

5. TRAFİK KAZA VERİLERİNİN ANALİZİ

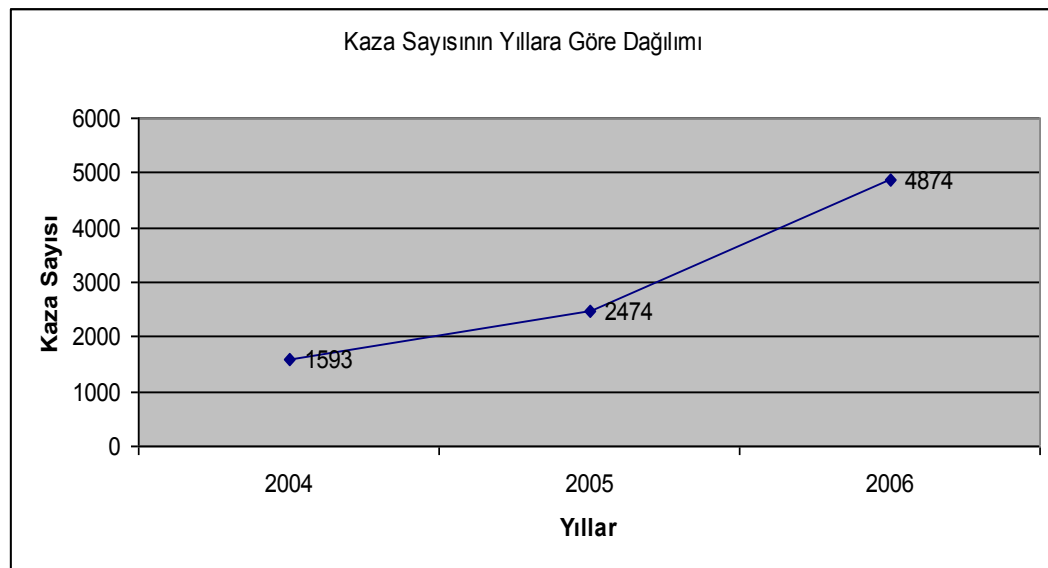
5.1 Genel İstatistikler

Kaza analiz çalışmaları, genellikle istatistiksel verilere dayalı çalışmalardır. Çünkü trafik kazaları, yeri ve zamanı önceden bilinmeyen, kameralara takılanlar hariç meydana gelişleri genellikle objektif olarak izlenemeyen olaylardır.

Denizli kenti 2004, 2005 ve 2006 yıllarına ait trafik kaza verisinin yıllara, aylara, haftanın günlerine ve saatlere göre dağılımı aşağıda tablolar halinde verilmiştir.

Tablo 5.1 Yıllara göre kaza dağılımı

Yıllara Göre Dağılım		
Yıllar	Kaza Sayısı	Sıklık
2004	1593	17,817
2005	2474	27,670
2006	4874	54,513
TOPLAM	8941	100,000

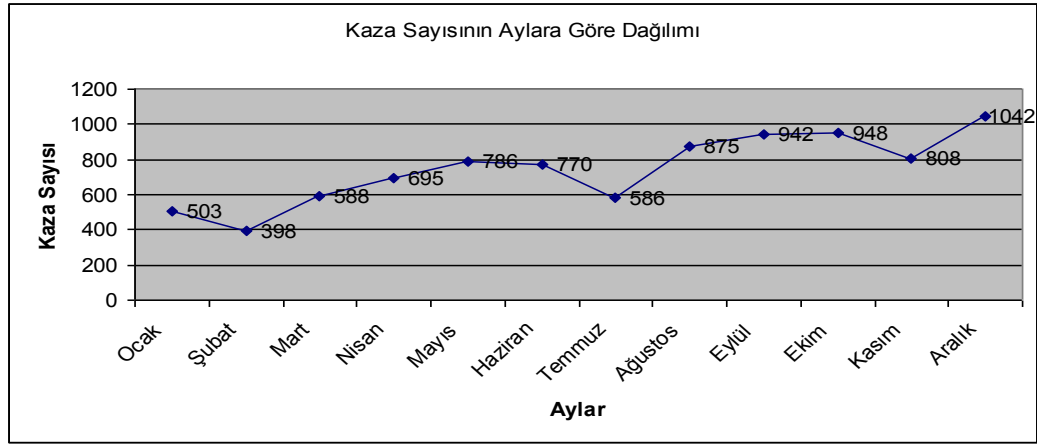


Şekil 5.1 Yıllara göre kaza dağılımı

Tablo 5.1' den, kaza sayısının 2005 yılında 2004 yılına oranla yaklaşık 1,5 kat, 2006 yılında ise 2005 yılına oranla yaklaşık 2 kat artmış olduğu çok açık bir şekilde görülmektedir.

Tablo 5.2 Aylara göre kaza dağılımı

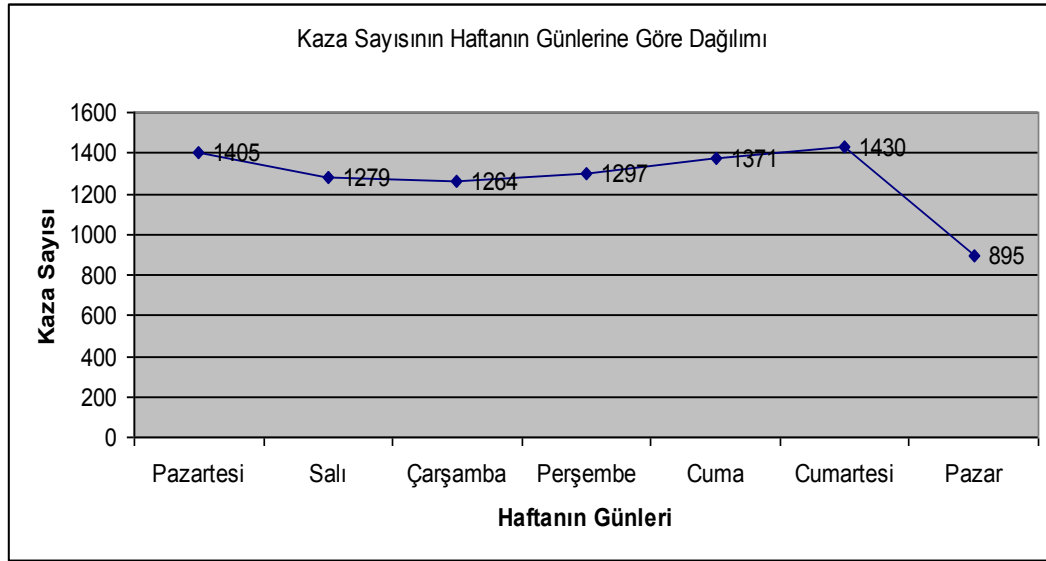
Aylara Göre Dağılım		
Aylar	Kaza Sayısı	Sıklık
Ocak	503	5,626
Şubat	398	4,451
Mart	588	6,576
Nisan	695	7,773
Mayıs	786	8,791
Haziran	770	8,612
Temmuz	586	6,554
Ağustos	875	9,786
Eylül	942	10,536
Ekim	948	10,603
Kasım	808	9,037
Aralık	1042	11,654
TOPLAM	8941	100,000

**Şekil 5.2** Aylara göre kaza dağılımı

Tablo 5.2 ve Şekil 5.2' den anlaşılacağı üzere, üç yıl içinde en fazla kaza aralık ayında görülürken, en az kaza ise şubat ayında görülmüştür. Kış ve bahar aylarında kaza sayılarında görülen artış, hava koşullarının düzelmesiyle yaz aylarında azalan bir eğilim göstermiştir.

Tablo 5.3 Haftanın günlerine göre kaza dağılımı

Haftanın Günlerine Göre Dağılım		
Günler	Kaza Sayısı	Sıklık
Pazartesi	1405	15,714
Salı	1279	14,305
Çarşamba	1264	14,137
Perşembe	1297	14,506
Cuma	1371	15,334
Cumartesi	1430	15,994
Pazar	895	10,010
TOPLAM	8941	100,00

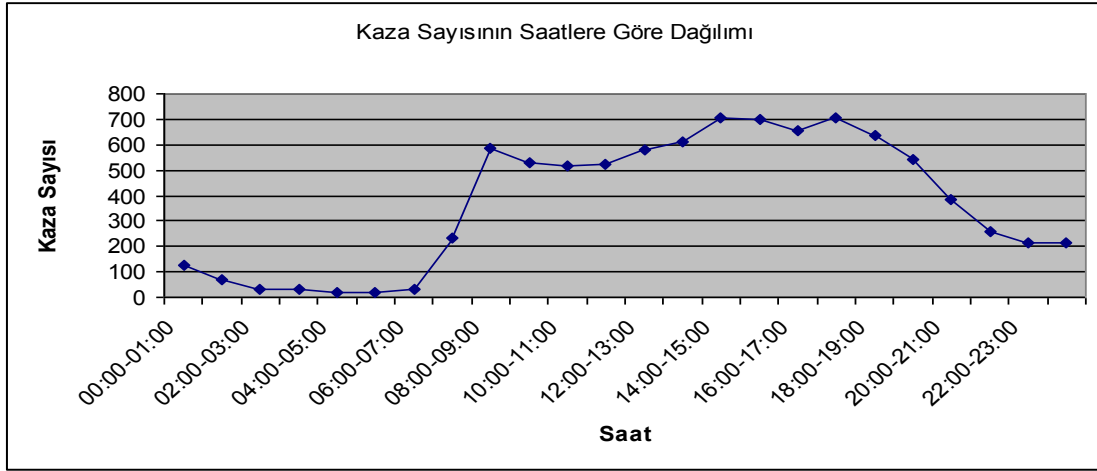


Şekil 5.3 Haftanın günlere göre kaza dağılımı

Üç yıl boyunca en çok kazanın, tatil olan ve trafiğin çok yoğun olduğu cumartesi günü meydana geldiği, bunu haftanın ilk iş günü olan pazartesi gününün izlediği, en az kazanın ise pazar günü meydana geldiği açıkça görülmektedir.

Tablo 5.4 Saatlere göre kaza dağılımı

Saatlere Göre Dağılım		
Saatler	Kaza Sayısı	Sıklık
00:00-01:00	123	1,380
01:00-02:00	68	0,763
02:00-03:00	29	0,325
03:00-04:00	30	0,337
04:00-05:00	19	0,213
05:00-06:00	17	0,191
06:00-07:00	32	0,359
07:00-08:00	234	2,626
08:00-09:00	587	6,587
09:00-10:00	531	5,958
10:00-11:00	516	5,790
11:00-12:00	522	5,857
12:00-13:00	579	6,497
13:00-14:00	610	6,845
14:00-15:00	705	7,911
15:00-16:00	700	7,855
16:00-17:00	654	7,338
17:00-18:00	704	7,899
18:00-19:00	639	7,170
19:00-20:00	539	6,048
20:00-21:00	387	4,342
21:00-22:00	261	2,929
22:00-23:00	214	2,401
23:00-24:00	212	2,379
TOPLAM	8912	100,000



Şekil 5.4 Saatlere göre kaza dağılımı

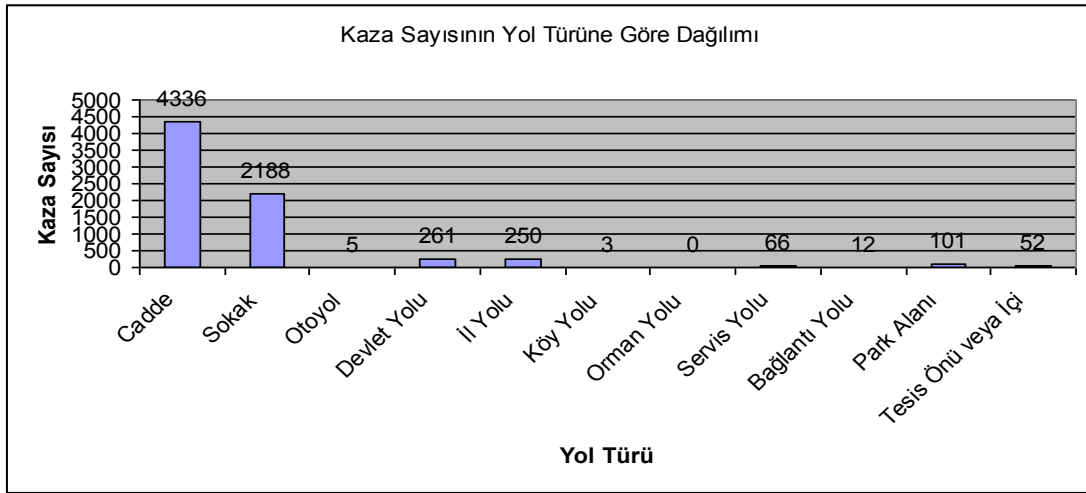
Şekil 5.4'den de görüldüğü üzere kaza sayıları trafiğin yoğun olduğu sabah ve akşam saatlerinde (pik saatlerde) maksimum seviyeye çıkmaktadır.

Bu verilerin kazaya etki eden yol türü, yol bölünme, oluşumuna göre kaza türü, hava durumu, gün durumu, yolun kaplama cinsi faktörlerine göre dağılımları ise şu şekildedir;

Tablo 5.5 Yol türüne göre kaza dağılımı

Yol Türü	Kaza Sayısı	Sıklık
Cadde	4336	59,610
Sokak	2188	30,080
Otoyol	5	0,069
Devlet Yolu	261	3,588
İl Yolu	250	3,437
Köy Yolu	3	0,041
Orman Yolu	0	0,000
Servis Yolu	66	0,907
Bağlantı Yolu	12	0,165
Park Alanı	101	1,389
Tesis Önü veya İçi	52	0,715
TOPLAM	7274	100,000

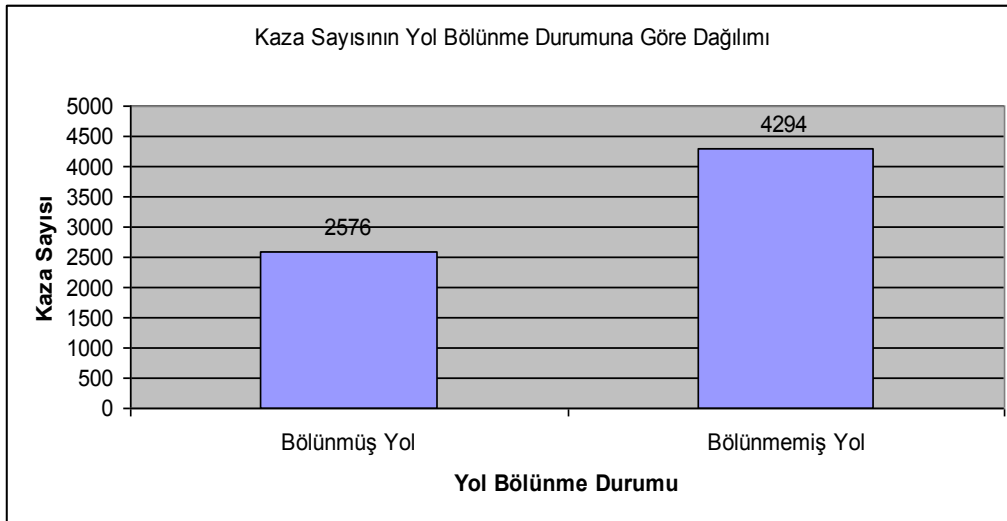
Şekil 5.5' de kazaların daha çok şehir merkezlerinde bulunan cadde ve sokaklarda meydana gelmiş olduğu açıkça görülmektedir.



Şekil 5.5 Yol türüne göre kaza dağılımı

Tablo 5.6 Yol bölünme durumuna göre kaza dağılımı

Yol Bölünme	Kaza Sayısı	Sıklık
Bölünmüş Yol	2576	37,496
Bölünmemiş Yol	4294	62,504
TOPLAM	6870	100,000

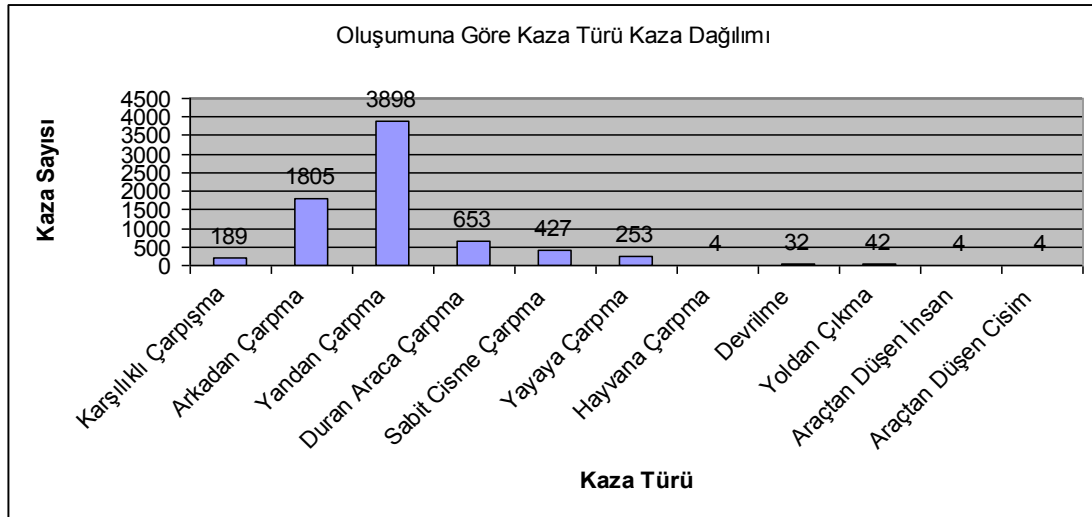


Şekil 5.6 Yol bölünme durumuna göre kaza dağılımı

Tablo 5.6' dan da anlaşılacağı üzere bölünmemiş yollarda meydana gelen kaza sayısı bölünmüş yollarda meydana gelen kaza sayısının yaklaşık iki katıdır.

Tablo 5.7 Oluşumuna göre kaza dağılımı

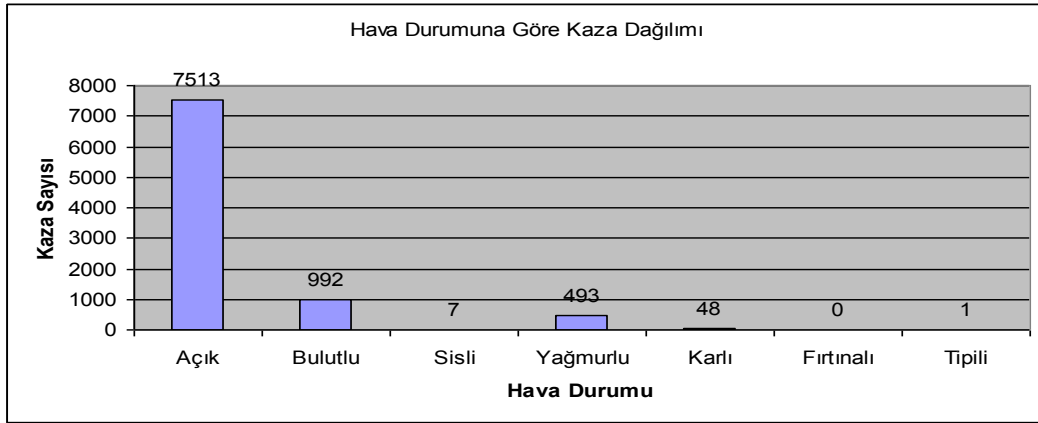
Oluşumuna Göre Kaza Türü	Kaza Sayısı	Sıklık
Karşılıklı Çarpışma	189	2,585
Arkadan Çarpma	1805	24,689
Yandan Çarpma veya Çarpışma	3898	53,317
Duran Araca Çarpma	653	8,932
Sabit Cisme Çarpma	427	5,841
Yayaya Çarpma	253	3,461
Hayvana Çarpma	4	0,055
Devrilme	32	0,438
Yoldan Çıkma	42	0,574
Araçtan Düşen İnsan	4	0,055
Araçtan Düşen Cisim	4	0,055
TOPLAM	7311	100,000

**Şekil 5.7** Oluşumuna göre kaza dağılımı

Şekil 5.7' den görüldüğü üzere yandan çarpma ve arkadan çarpma tüm kaza türleri arasında önemli bir orana sahiptir. Araçtan düşen insan, araçtan düşen cisim ve hayvana çarpma türü kazalar ise en az rastlanan kaza türleri arasındadır.

Tablo 5.8 Hava durumuna göre kaza dağılımı

Hava Durumu	Kaza Sayısı	Sıklık
Açık	7513	82,980
Bulutlu	992	10,956
Sisli	7	0,077
Yağmurlu	493	5,445
Karlı	48	0,530
Fırtınalı	0	0,000
Tipili	1	0,011
TOPLAM	9054	100,000

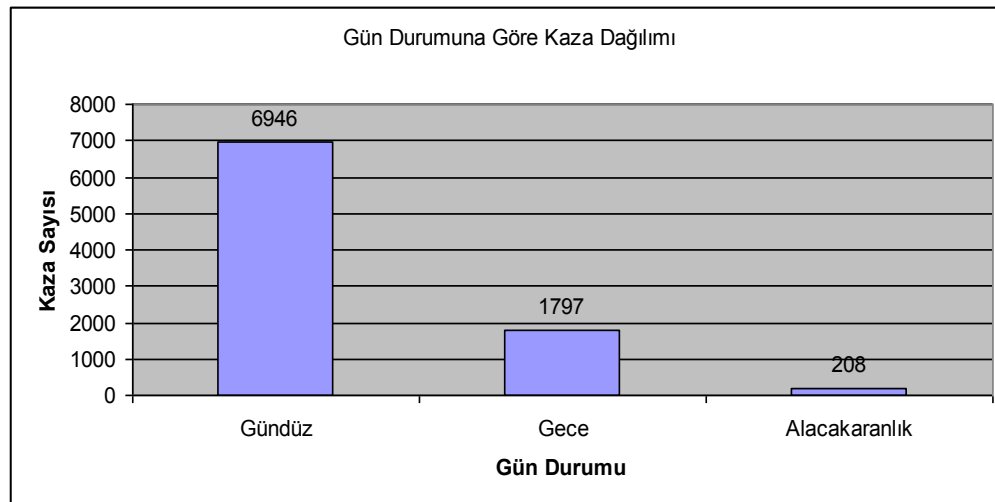


Şekil 5.8 Hava durumuna göre kaza dağılımı

Kazalar daha çok açık havalarda meydana gelmektedir. Bunun sebebi ise sürücülerin olumlu hava şartlarında yüksek hız yapmaları ve daha dikkatsiz olmalarıdır.

Tablo 5.9 Gün durumuna göre kaza dağılımı

Gün Durumu	Kaza Sayısı	Sıklık
Gündüz	6946	77,600
Gece	1797	20,076
Alacakaranlık	208	2,324
TOPLAM	8951	100,000

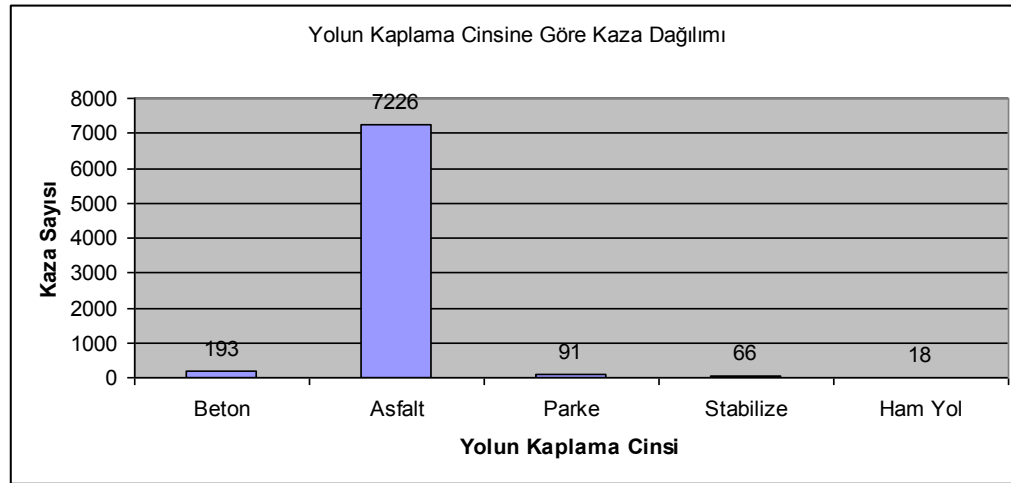


Şekil 5.9 Gün durumuna göre kaza dağılımı

Görüldüğü üzere gündüz meydana gelen kazaların sayısı gece meydana gelen kazaların sayısının oldukça üstündedir.

Tablo 5.10 Yolun kaplama cinsine göre kaza dağılımı

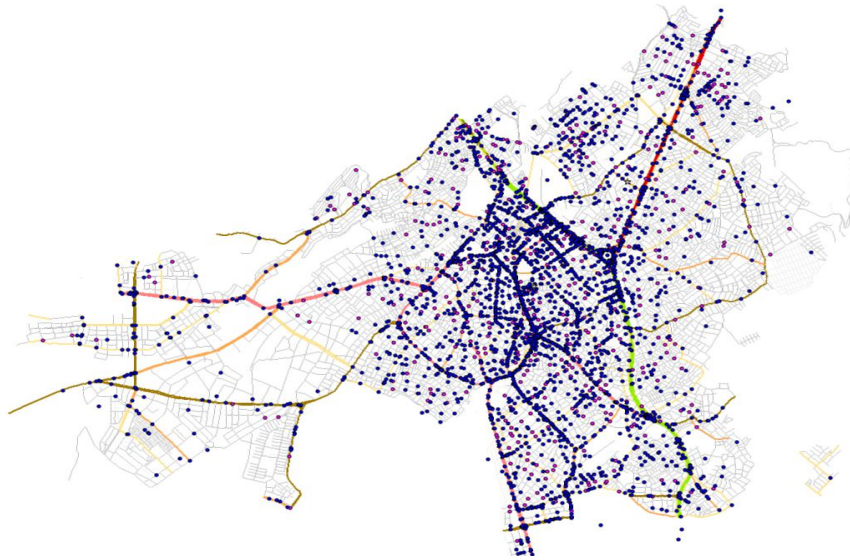
Yolun Kaplama Cinsi	Kaza Sayısı	Sıklık
Beton	193	2,541
Asfalt	7226	95,154
Parke	91	1,198
Stabilize	66	0,869
Ham Yol	18	0,237
TOPLAM	7594	100,000

**Şekil 5.10** Yolun kaplama cinsine göre kaza dağılımı

Denizli kentiçi yolları genel olarak asfalt kaplamalı yollar olduğu için en fazla kaza bu tür yollarda meydana gelmiştir. Bunu çok daha geriden beton ve parke kaplamalı yollar izlemektedir.

5.2 Konum (x, y) Verisinin K-Ortalamlar ile Kümeleme Analizi

Konum verisi, kazanın meydana geldiği noktanın, MAPINFO yazılımında bulunan sokak tanımlama sisteminin kullanılmasıyla elde edilen veridir. Bütün detayları içeren trafik kaza verisi, Coğrafi Bilgi Sistemi (CBS) veritabanında kaydedilmiştir. Coğrafi Bilgi Sistemi (CBS), ülke veya yeryüzü referans sistemine bağlı yer-referanslı (geo-referenced data) verileri girme, arşivleme, analiz etme ve görüntüyle çıktı elde etmek için geliştirilmiş bilgisayar tabanlı bir sistemdir (Yıldırım 1994). Şekil 5.11 Denizli şehri GIS haritasındaki örnek yönlendirilmiş veriyi göstermektedir.



Şekil 5.11 GIS’ de Denizli haritasındaki örnek yerleştirilmiş trafik kaza verisi

Konum verisinin k-ortalamlar kümeleme analizinde optimum küme sayısı bilinmediği için başlangıç küme sayısı tahmini olarak 15 seçilmiştir. Böylelikle elde edilen minimum iterasyonlar incelendiğinde;

Tablo 5.11 Konum verisinin k-ortalamlar kümeleme analizi sonucu minimum iterasyonu

Minimum İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgi Grafiği
2	2	65,36	
4	3	43,46	
9	4	31,43	
10	5	24,45	
13	6	18,81	
16	7	15,56	
20	8	13,56	
22	9	12,2	
27	10	11,02	
29	11	9,54	
31	12	8,97	
36	13	7,92	
38	14	7,22	
40	15	6,82	

Optimum küme sayısı olarak, değişim yüzdesinde belirgin bir düşüşün görülmediği küme sayısı olan 7 seçilebilir. Bu seçilen optimum küme sayısına göre konum verisi tekrar analiz edilirse;

Tablo 5.12 Konum verisinin k-ortalamlar kümeleme analizi sonucu iterasyonu

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	74,33	
2	2	65,36	
3	2	65,36	
4	3	43,46	
5	3	43,59	
6	3	43,59	
7	4	32,6	
8	4	32,6	
9	4	32,48	
10	5	24,45	
11	5	25,16	
12	5	25,82	
13	6	18,81	
14	6	18,81	
15	6	18,81	
16	7	15,56	
17	7	15,56	
18	7	15,56	

NCSS’de her bir küme, random starts kısmında girilen sayı kadar iterasyona tabi tutulur. İterasyon tablosu, bu kümelere ait iterasyonları ve değişim yüzdelere verir. Elde edilen küme merkezleri ise Tablo 5.13’ deki gibidir.

Tablo 5.13 Konum verisinin k-ortalamlar kümeleme analizi sonucu küme merkezleri

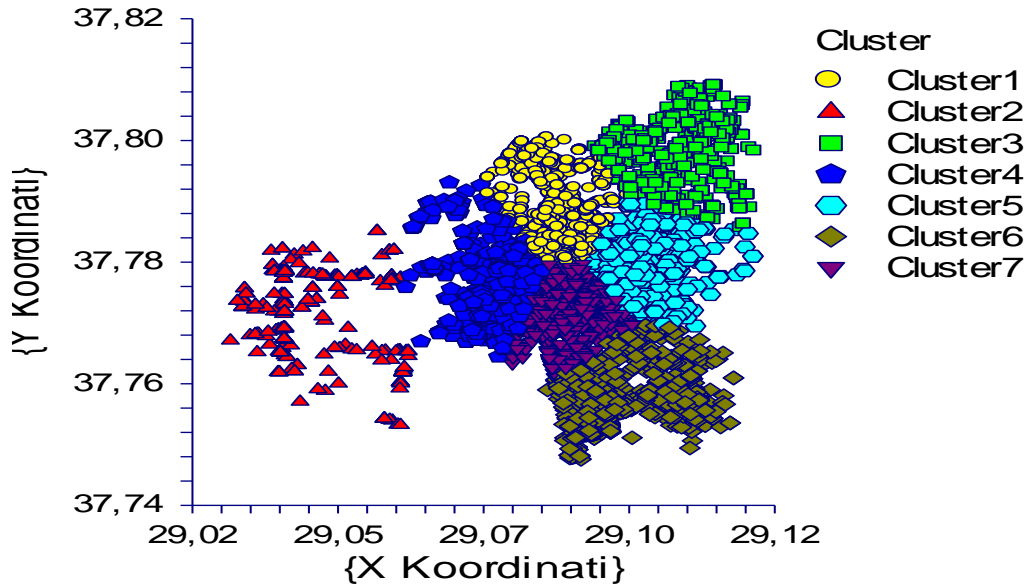
Küme No	Değişkenler		Veri Sayısı
	x	y	
1	29,084884	37,787184	2480
2	29,041337	37,772047	302
3	29,102469	37,798443	973
4	29,071570	37,777544	835
5	29,096437	37,780394	1589
6	29,094854	37,759552	1274
7	29,085393	37,772304	1645

Elde edilen tüm konum verisine ait uzaklıklar matrisinin boyutu çok büyük olduğu için, Tablo 5.14’ de sadece 10 elemana ait uzaklık matrisi örnek olarak gösterilmiştir.

Tablo 5.14 Konum verisinin k-ortalamlar kümeleme analizi sonucu uzaklıklar matrisi

Uzaklıklar Matrisi								
Sıra	Küme	Uzaklık1	Uzaklık2	Uzaklık3	Uzaklık4	Uzaklık5	Uzaklık6	Uzaklık7
1	6	2,0881	3,6918	3,211	1,7961	1,6449	0,5748	0,8612
2	5	1,0082	4,4046	1,3439	2,0275	0,2272	1,9605	1,2735
3	3	2,0863	5,6177	0,4475	3,3893	1,9861	3,6739	2,9819
4	7	1,8124	2,7705	3,2993	0,985	1,8962	1,5175	0,8113
5	1	0,0738	3,5649	1,6572	1,3025	1,1259	2,4909	1,2961
6	1	0,2271	3,7723	1,5167	1,4466	0,8389	2,2988	1,1852
7	3	1,3991	4,9876	0,4819	2,6653	1,1069	2,8068	2,1237
8	2	4,0068	0,4384	5,6096	2,8114	4,7441	4,682	3,8387
9	5	1,0936	3,9922	1,9838	1,6461	0,438	1,401	0,6651
10	1	0,1798	3,5293	1,7573	1,2005	0,9716	2,2525	1,0577

Konum verisinin kümeler dağılımı ise Şekil 5.12'deki gibi elde edilmiştir.

**Şekil 5.12** Konum verisinin k-ortalamlar küme analizi sonucu küme grafiği

Bu analizde küme merkezlerine göre elde edilen kara noktalar ise Tablo 5.15' de gösterilmiştir.

Tablo 5.15 Konum verisinin k-ortalamlar kümeleme analizi sonucu belirlenen kara noktalar

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
İstasyon Kavşağı	1	2480
Çınar Mey	7	1645
Deliktaş Kavşağı (Emniyet Müd.)	5	1589
Fatih Kavşağı	6	1274
Ankara Bulvarı (Sevindik Alt Geçidi)	3	973
Özay Gönülüm Meydanı	4	835
Yeni Adliye	2	302

5.3 Konum (x, y) Verisinin Bulanık Kümeleme Analizi

Daha önce de bahsedildiği üzere bulanık kümeleme metodunda bir elemanın bir kümeye aitlik derecesi (üyelik değeri) vardır. Bu üyelik değeri 0 ile 1 arasında sonsuz değer alabilmektedir ve bu üyelik değerlerinin toplamı daima 1'e eşittir. Bundan dolayı eleman **en yüksek üyelik değerine** sahip olduğu kümeye atanır.

Burada konum verisinin bulanık kümeleme analizi için ilk olarak NCSS programı kullanılmış, fakat büyük veri boyutu nedeniyle analizin çok uzun sürmesinden ve kümelemenin grafiğinin çizilememesinden dolayı NCSS programının kullanımından vazgeçilerek analize MATLAB programı ile devam edilmiştir.

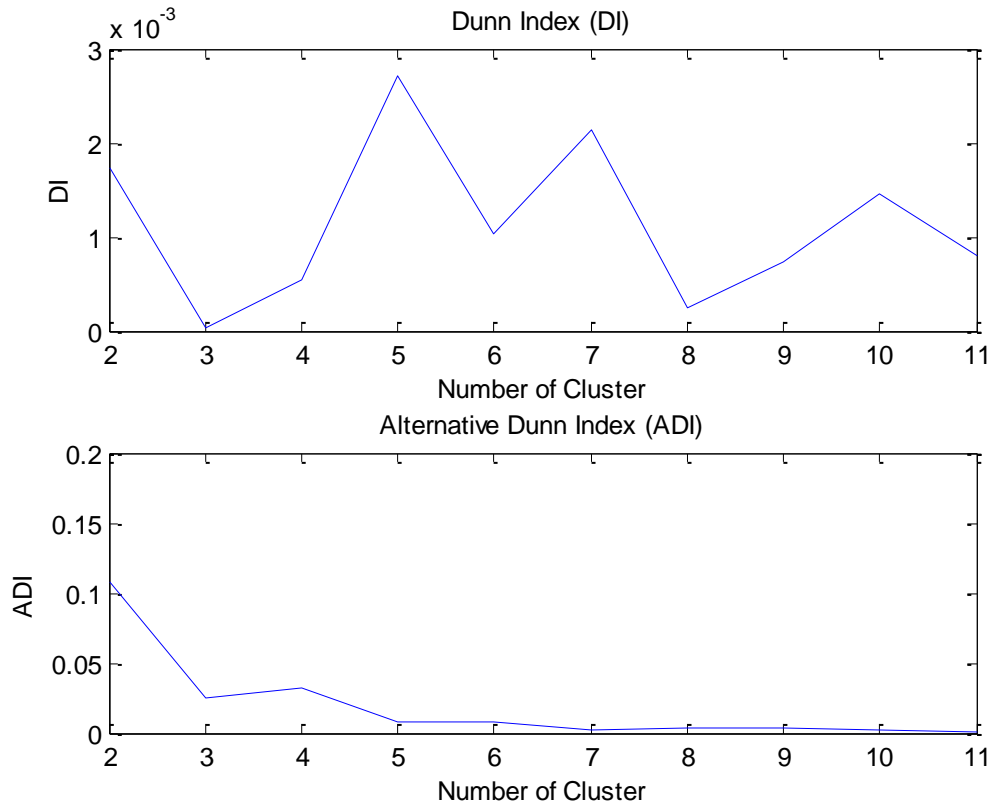
Burada da optimum küme sayısı bilinmediğinden dolayı, veriler iki kere analize tabi tutulmuştur. İlk analizde, elde edilen *ayırma katsayısı (partition coefficient, PC)*, *sınıflandırma entropisi (classification entropy, CE)*, *ayırma indeksi (partition index, SC)*, *ayırıştırma indeksi (separation index, S)*, *Xie ve Beni indeksi (Xie and Beni's Index, XB)*, *Dunn indeksi (Dunn's index, DI)* değerlerine bakılarak optimum küme sayısına karar verilmiştir. Fakat ayırma katsayısının maksimum, sınıflandırma entropisinin minimum değerinin görüldüğü küme sayısının genelde iki veya üç olmasının mantıksız olmasından dolayı ve ayırma katsayısının, verilerin özelliklerine direk bağlantısının eksik olmasının bir dezavantaj olmasından dolayı, bu iki katsayının değerlendirilmesinden vazgeçilmiştir.

Optimum küme sayısı ise SC, S ve XB' in minimum değerlerine denk gelen 11 olarak seçilmiştir.

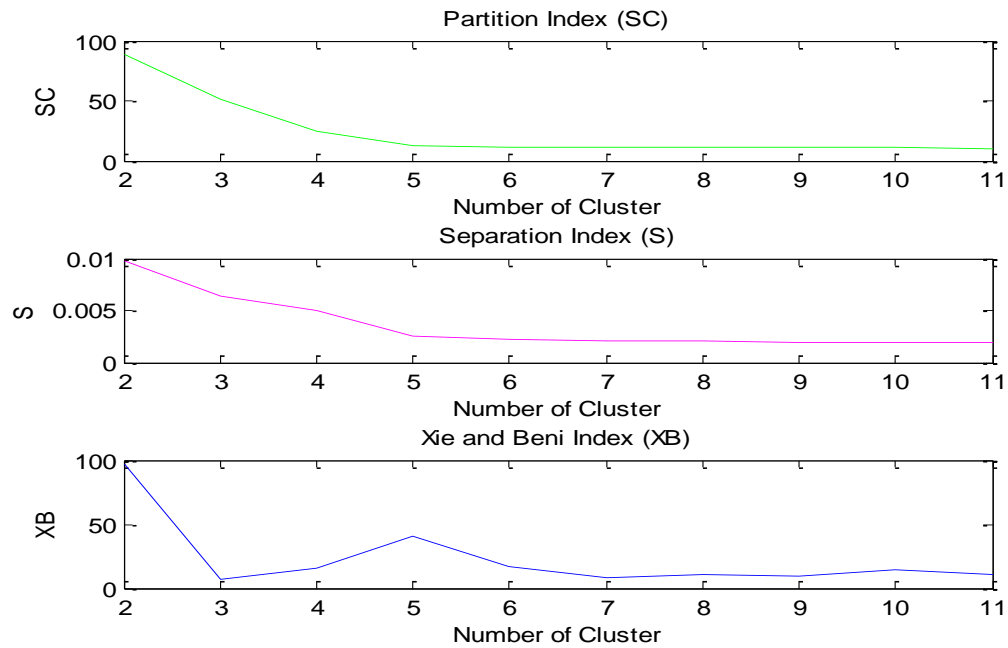
Tablo 5.16 Doğrulama katsayıları

Küme No	SC	S	XB	DI
2	88,6370	0,0097	96,548	0,0017342
3	51,2240	0,0064	10,879	2,24E-05
4	24,6720	0,0050	16,086	0,00054551
5	12,5210	0,0025	40,156	0,0027215
6	11,7100	0,0022	17,276	0,0010383
7	11,6490	0,0020	14,228	0,002147
8	11,1400	0,0021	10,769	0,00025443
9	10,9860	0,0020	11,737	0,00073721
10	10,8830	0,0020	14,86	0,0014641
11	10,6270	0,0019	10,762	0,00079156

Buna göre MATLAB' da maksimum küme sayısı 11 olarak girildiğinde elde edilen sonuçlar;



Şekil 5.13 Dunn indeksi (DI) ve alternatif Dunn indeksi (ADI)

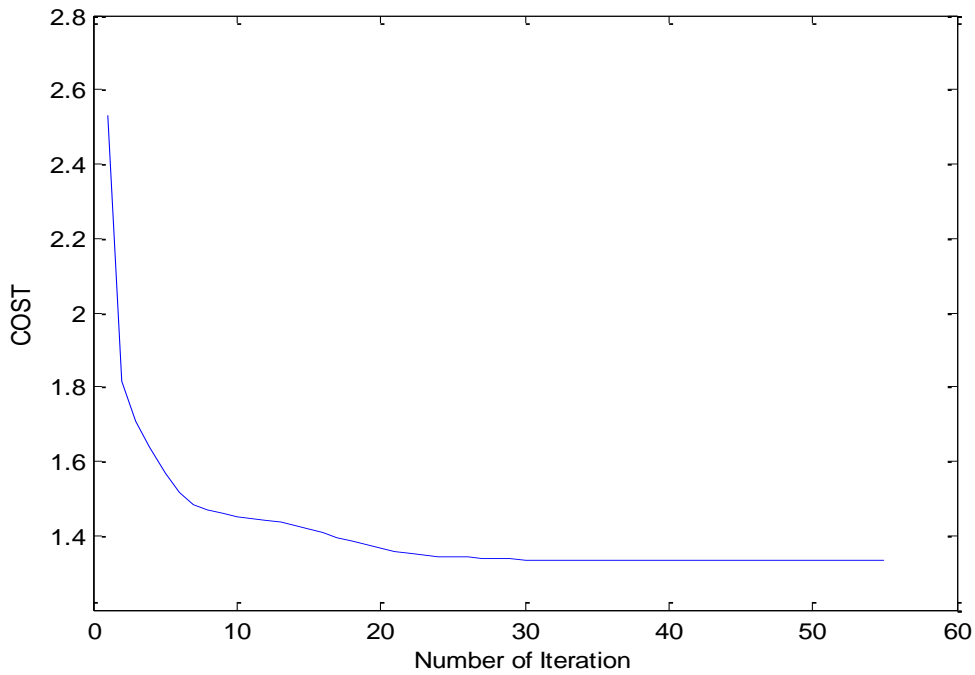


Şekil 5.14 Ayırma indeksi (PC), ayırıştırma indeksi (S), Xie ve Beni indeksi (XB)

Amaç fonksiyonun iterasyon adımları Tablo 5.17' deki gibidir;

Tablo 5.17 Amaç fonksiyonu iterasyon adımları

İt. No	Amaç Fonk.	İt. No	Amaç Fonk.	İt. No	Amaç Fonk.	İt. No	Amaç Fonk.	İt. No	Amaç Fonk.
1	2,5314	12	1,4416	23	1,3477	34	1,3339	45	1,3333
2	1,8145	13	1,4358	24	1,3447	35	1,3337	46	1,3333
3	1,708	14	1,4278	25	1,3424	36	1,3336	47	1,3333
4	1,6376	15	1,4176	26	1,3406	37	1,3335	48	1,3333
5	1,5686	16	1,4064	27	1,339	38	1,3334	49	1,3333
6	1,5145	17	1,3956	28	1,3376	39	1,3334	50	1,3333
7	1,484	18	1,3855	29	1,3365	40	1,3334	51	1,3333
8	1,4675	19	1,3759	30	1,3356	41	1,3334	52	1,3333
9	1,4578	20	1,3665	31	1,3349	42	1,3334	53	1,3333
10	1,4513	21	1,3583	32	1,3345	43	1,3333	54	1,3333
11	1,4464	22	1,352	33	1,3341	44	1,3333	55	1,3333



Şekil 5.15 Amaç fonksiyonu grafiği

Konum veri setinin çok büyük olmasından dolayı her bir elemanın üyelik derecesini göstermek yerine, örnek olarak 5 elemanın üyelik dereceleri Tablo 5.18' de gösterilmiştir. Elemanlar yüksek üyelik değerine sahip oldukları kümeye aittirler. Sarı hücreler hangi elemanın hangi kümeye ait olduğunu göstermektedir.

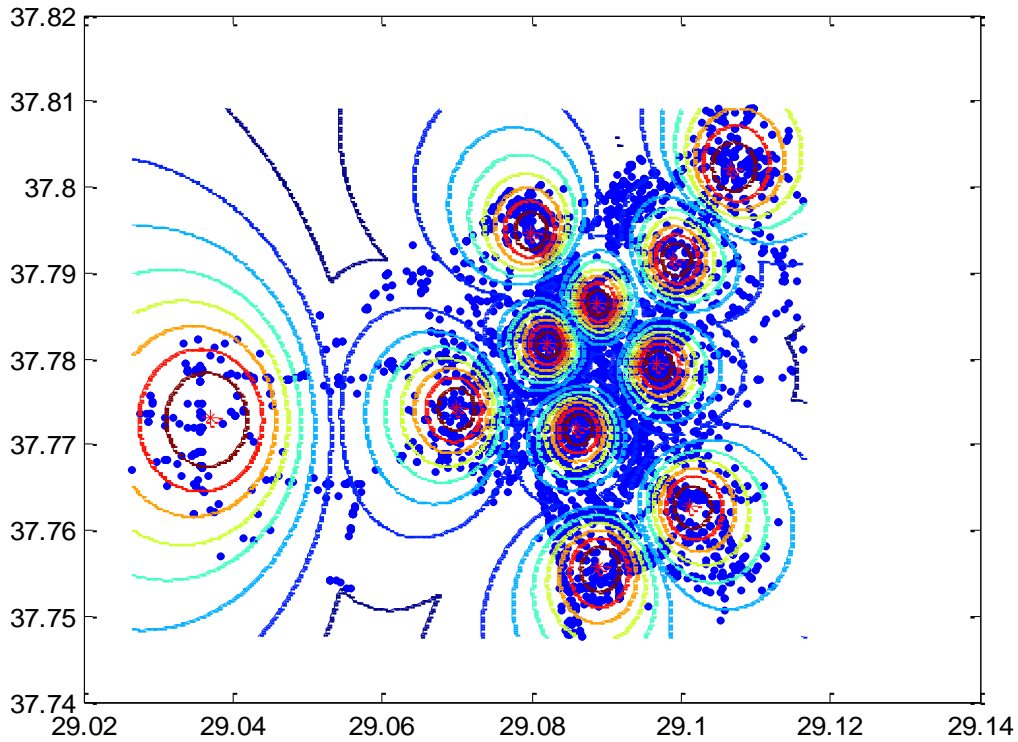
Tablo 5.18 Eleman üyelik dereceleri

Küme No	Eleman No				
	1	2	3	4	5
1	0,047315	0,037842	0,00019467	0,11425	0,22417
2	0,041651	0,010648	9,3789e-005	0,27779	0,024679
3	0,03314	0,10387	0,00035198	0,053878	0,44852
4	0,20156	0,035137	0,00014765	0,25527	0,037229
5	0,053094	0,62485	0,00031696	0,051701	0,042116
6	0,010114	0,019192	0,99724	0,013447	0,013664
7	0,017486	0,019957	0,00026063	0,036372	0,14466
8	0,0071925	0,0023314	3,4913e-005	0,017696	0,0038993
9	0,019971	0,11484	0,001154	0,025915	0,041368
10	0,45523	0,010736	8,0098e-005	0,10625	0,0091146
11	0,11325	0,020602	0,00012466	0,047427	0,010577

Elde edilen kümelerin merkezleri ise Tablo 5.19' daki gibidir.

Tablo 5.19 Konum verisinin bulanık kümeleme analizi sonucu küme merkezleri

Küme No	Küme Merkezleri	
	x	y
1	29,082	37,782
2	29,07	37,774
3	29,089	37,787
4	29,086	37,772
5	29,097	37,779
6	29,107	37,802
7	29,08	37,795
8	29,037	37,773
9	29,099	37,792
10	29,089	37,756
11	29,101	37,763



Şekil 5.16 Konum verisinin bulanık kümeleme analizi sonucu küme grafiği

Denizli kentine ait trafik kaza verisinin bulanık kümeleme analizi sonucundaki küme merkezlerine göre elde edilen kara noktalar ise Tablo 5.20' deki gibidir.

Tablo 5.20 Konum verisinin bulanık kümeleme sonucu belirlenen kara noktalar

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
İstasyon Kavşağı	3	1389
Çınar Meydanı	4	1349
Vilayet Önü Kavşağı	1	1303
Deliktaş Kavşağı	5	1253
Demokrasi Meydanı	10	662
Özay Gönüm Meydanı	2	648
Fatih Kavşağı	11	631
25. Sokak	9	605
İzmir Bulvarı (Zafer Katrancı Köprülü Kav.)	7	509
Sevindik	6	501
Yeni Adliye	8	248
	TOPLAM	9098

5.4 K-Ortalamalar Metodu ile Diğer Kümeleme Analizleri

Trafik kazalarının oluşmasında insan faktörü başta olmak üzere birden fazla faktörün etkisi bulunmaktadır. Bunun için, bu faktörlerin de trafik kazaları üzerindeki etkisinin araştırılması ve bu konuda alınabilecek önlemlerin belirlenmesi gerekmektedir. Bu çalışmada trafik kazalarına etki eden diğer faktörler kazanın yeri, kazanın türü, hava-gün durumu ve yol-çevre özellikleri olarak ele alınmış ve bu faktörlere göre de ayrıca kümeleme analizi yapılmıştır. Burada kazaların genel olarak nerede, hangi türde, hangi koşullarda meydana geldiğinin belirlenmesi amaçlanmıştır.

Kümeleme analizine başlamadan önce ilk olarak, Denizli kenti için 2004, 2005 ve 2006 yıllarına ait trafik kaza verilerinin hangi ikili parametrelere göre kümeleme analizine tabi tutulacağı belirlenmesi gerekmektedir. Bunun için, anlamlılık düzeyi yüksek olan ikili parametreleri seçmek daha mantıklı olacaktır.

Kümeleme analizinde kullanılacak parametreler;

Yol Türü: Cadde, sokak, otoyol, devlet yolu, il yolu, köy yolu, orman yolu

Yol Bölünme: Bölünmüş yol, bölünmemiş yol

Yerleşim Durumu: Yerleşim yeri, yerleşim yeri dışı

Oluşumuna Göre Kaza Türü: Karşılıklı çarpışma, arkadan çarpma, yandan çarpma veya yandan çarpışma, duran araca çarpma, sabit cisme çarpma, yayaya çarpma, hayvana çarpma, devrilme, yoldan çıkma, araçtan düşen insan, araçtan düşen cisim

Hava Durumu: Açık, bulutlu, sisli, yağmurlu, karlı, fırtınalı, tipili

Gün Durumu: Gündüz, gece, alacakaranlık

Yolda Yön: Tek yönlü, iki yönlü

Yolun Kaplama Cinsi: Beton, asfalt, parke, stabilize, ham yol

Bu durumda örneğin yol türünün cadde, sokak, otoyol v.b olması halinde yolun bölünmüş olup olmadığındaki (yol bölünme) kaza sayıları, yerleşim yeri veya yerleşim yeri dışı olması (yerleşim durumu) halindeki kaza sayıları, havanın açık, bulutlu, sisli v.b (hava durumu) olması halindeki kaza sayıları, gün durumunun gündüz, gece, alacakaranlık olması halindeki kaza sayıları, yolun tek veya iki yönlü (yolda yön) olması halindeki kaza sayıları, yol kaplamasının beton, asfalt v.b (yolun kaplama cinsi) olması halindeki kaza sayıları ile, yine yol türünün cadde, sokak, otoyol v.b olması halinde, karşılıklı çarpışma, arkadan çarpma v.b (Oluşumuna göre kaza türü) kaza sayıları ile farklı kümelemeler yapılabilir. Bu iki parametrelili kümeleme kombinasyonları bir tablo halinde verilecek olursak,

Tablo 5.21 Kümeleme parametreleri

1. Parametre	2. Parametre
Yol Türü	Yol Bölünme
	Oluşumuna Göre Kaza Türü
	Hava Durumu
	Gün Durumu
	Yolda Yön
	Yolun Kaplama Cinsi
Yol Bölünme	Oluşumuna Göre Kaza Türü
	Hava Durumu
	Gün Durumu
	Yolun Kaplama Cinsi
Yerleşim Durumu	Oluşumuna Göre Kaza Türü
	Hava Durumu
	Gün Durumu
	Yolda Yön
	Yolun Kaplama Cinsi
Hava Durumu	Oluşumuna Göre Kaza Türü
	Yolun Kaplama Cinsi

5.4.1 Yol türü – yol bölünme

Daha öncede bahsedildiği üzere k-ortalamlar küme analizinde ilk etapta optimum küme sayısı bilinmediği için, veri setinin iki kere analize tabi tutulması gerekir. İlk analizde, maksimum küme sayısına rasgele bir küme sayısı girilir ve elde edilen analiz

çıktısı incelenir. Elde edilen çıktıda ilk olarak minimum iterasyon kısmında yer alan değişim yüzdesine bakılır. Burada değişimin çok açık bir şekilde değişmediği küme sayısı optimum küme sayısı olarak seçilir. İkinci analizde ise, maksimum küme sayısı ve raporlanacak küme sayısına bu elde edilen optimum küme sayısı girilerek analiz sonuçlandırılır.

Belirlenen optimum küme sayısı: 3

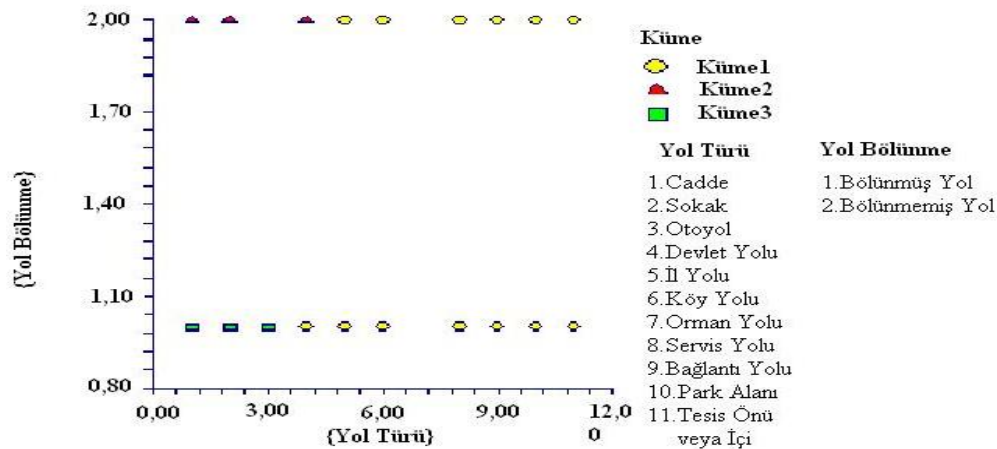
Tablo 5.22 Yol türü–yol bölünme, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	69,33	
2	2	49,94	
3	2	77,72	
4	3	55,01	
5	3	17,23	
6	3	42,69	

Tablo 5.23 Yol türü–yol bölünme küme merkezleri

Küme Merkez Değişkenleri	Küme1		
	Küme1	Küme2	Küme3
y	1,20846395	2	1
x	5,764890282	1,470816178	1,064737616
Veri Sayısı	638	4129	2039

Şekil 5.17 ve Tablo 5.23’ den, en çok kaza verisini içeren 2 numaralı kümeyi, bölünmemiş cadde ve sokakların oluşturduğu, bunu daha çok bölünmüş caddelerin oluşturduğu 3 numaralı kümenin takip ettiği ve son olarak da en az kaza verisini içeren 1 numaralı kümeyi, bölünmüş il ve köy yolarının oluşturduğu açıkça görülmektedir.



Şekil 5.17 Yol türü–yol bölünme küme grafiği

5.4.2 Yol türü – oluşumuna göre kaza türü

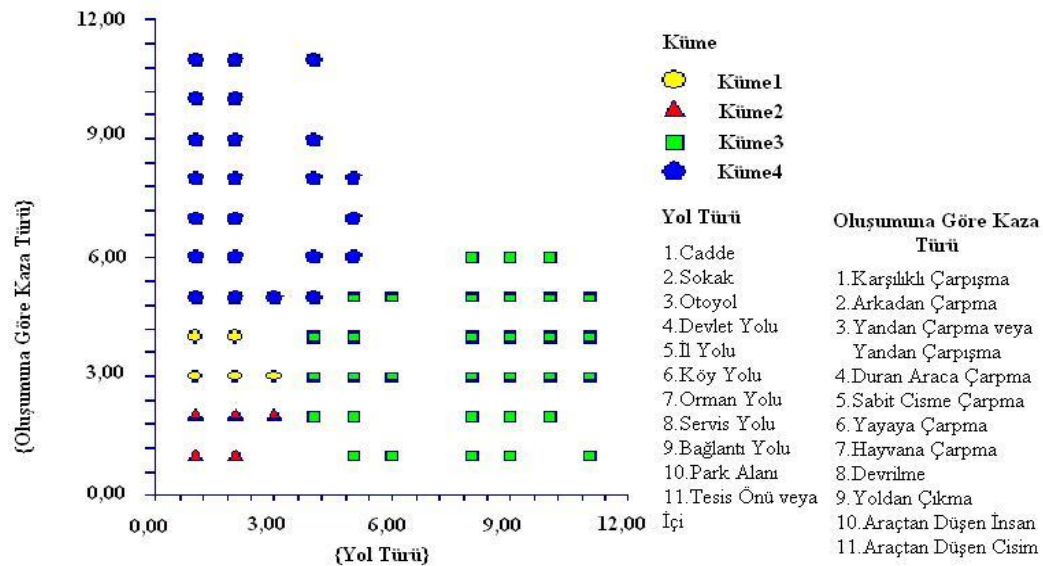
Optimum küme sayısı: 4

Tablo 5.24 Yol türü-oluşumuna göre kaza türü, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	65,12	
2	2	64,18	
3	2	65,28	
4	3	36,46	
5	3	36,67	
6	3	36,56	
7	4	30,3	
8	4	26,59	
9	4	26,59	

Tablo 5.25 Yol türü-oluşumuna göre kaza türü, küme merkezleri

Küme Merkez Değişkenleri	Kümeler			
	Küme1	Küme2	Küme3	Küme4
y	3,143033292	1,897583429	3,043227666	5,833587786
x	1,413316893	1,150172612	6,200288184	1,551145038
Veri Sayısı	4055	1738	694	655



Şekil 5.18 Yol türü-oluşumuna göre kaza türü, küme grafiği

Küme 1, cadde ve sokaklarda daha çok yandan çarpma veya duran araca çarpma, küme 2, cadde ve sokaklarda daha çok karşılıklı çarpma ve arkadan çarpma, küme 3, köy yollarında daha çok yandan çarpma ve duran araca çarpma, küme 4 ise cadde ve sokaklarda daha çok sabit cisme ve yayaya çarpma şeklinde kazaların meydana geldiği

göstermektedir. En çok kazanın ise, cadde ve sokaklarda yandan çarpma veya duran araca çarpma şeklinde meydana geldiği görülmektedir.

5.4.3 Yol türü – hava durumu

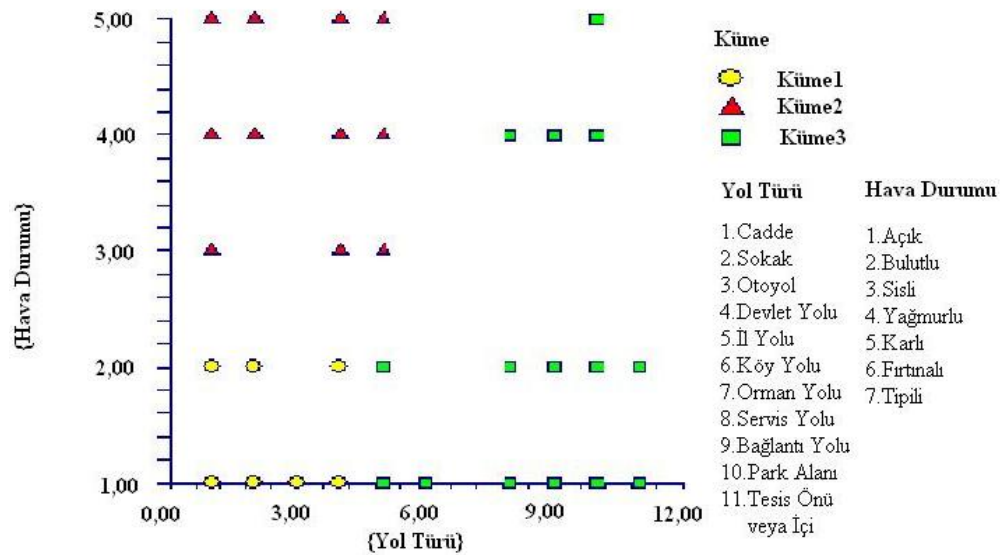
Optimum küme sayısı: 3

Tablo 5.26 Yol türü-hava durumu, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	65,69	
2	2	64,32	
3	2	80,51	
4	3	55,43	
5	3	23,57	
6	3	50,45	

Tablo 5.27 Yol türü-hava durumu, küme merkezleri

Küme Merkez Değişkenleri	Kümelere		
	Küme1	Küme2	Küme3
y	1,1119545	4,0842333	1,1938998
x	1,4422834	1,5809935	7,3224401
Veri Sayısı	6324	463	459



Şekil 5.19 Yol türü-hava durumu, küme grafiği

Burada 1 ve 2 numaralı küme kazaların daha çok cadde ve sokaklarda açık, yağmurlu ve karlı havalarda, 3 numaralı küme ise kazaların daha çok köy ve servis yollarında açık havalarda meydana geldiğini göstermektedir. En çok kazanın ise cadde, sokak, otoyol ve devlet yollarında açık havada meydana geldiği açıkça görülmektedir.

5.4.4 Yol türü – gün durumu

Optimum küme sayısı: 3

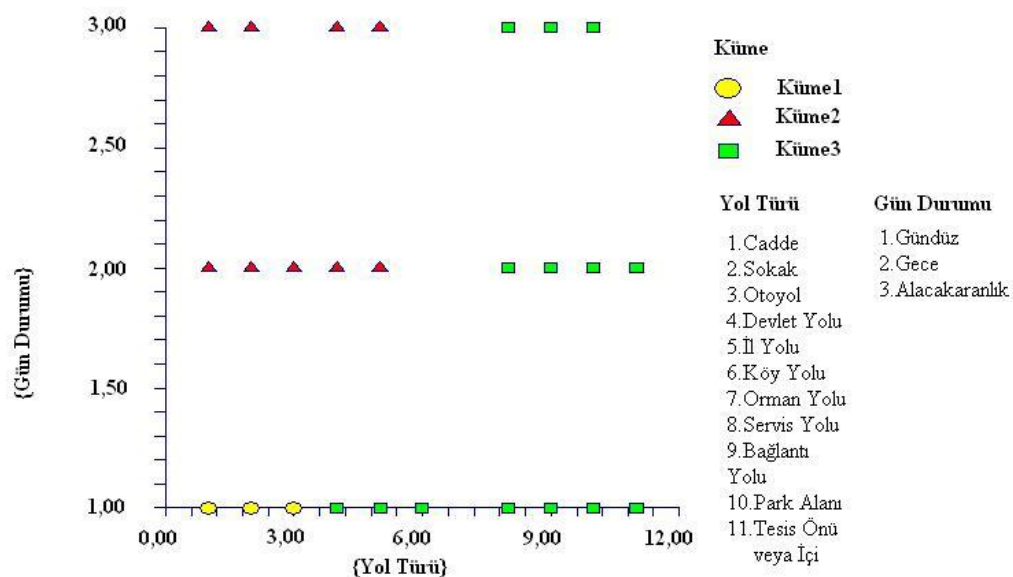
Tablo 5.28 Yol türü-gün durumu, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	65,72	
2	2	79,26	
3	2	54,53	
4	3	54,43	
5	3	22,61	
6	3	49,98	

Tablo 5.29 Yol türü-gün durumu, küme merkezleri

Küme Merkez Değişkenleri	Kümelere		
	Küme1	Küme2	Küme3
y	1	2,103318723	1,079173838
x	1,352208835	1,58672511	6,490533563
Veri Sayısı	4980	1597	581

Küme 1 ve küme 2, kazaların daha çok cadde ve sokaklarda gündüzleri ve geceleri, küme 2 daha çok otoyol ve devlet yollarında geceleri, küme 3 ise daha çok köy ve servis yollarında gündüzleri meydana geldiğini göstermektedir.



Şekil 5.20 Yol türü-gün durumu, küme grafiği

5.4.5 Yol türü – yolda yön

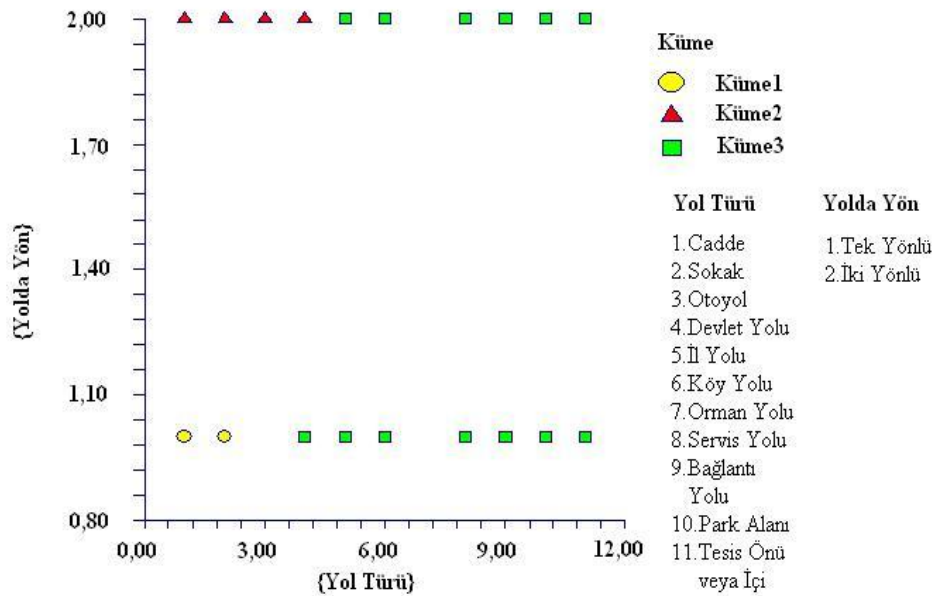
Optimum küme sayısı: 3

Tablo 5.30 Yol türü-yolda yön, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	66,81	
2	2	49,96	
3	2	49,96	
4	3	53,9	
5	3	40,53	
6	3	19,42	

Tablo 5.31 Yol türü-yolda yön, küme merkezleri

Küme Merkez Değişkenleri	Kümeler		
	Küme1	Küme2	Küme3
y	1	2	1,3678373
x	1,2460533	1,4614949	6,2162662
Veri Sayısı	1837	3532	541



Şekil 5.21 Yol türü-yolda yön, küme grafiği

Küme 1, tek yönlü cadde ve sokaklarda, küme 2, iki yönlü cadde ve sokaklarda, küme 3 ise, tek yönlü köy ve servis yollarında kazaların yoğunlaştığını göstermektedir. En çok kaza ise iki yönlü cadde ve sokaklarda meydana gelmektedir.

5.4.6 Yol türü – yolun kaplama cinsi

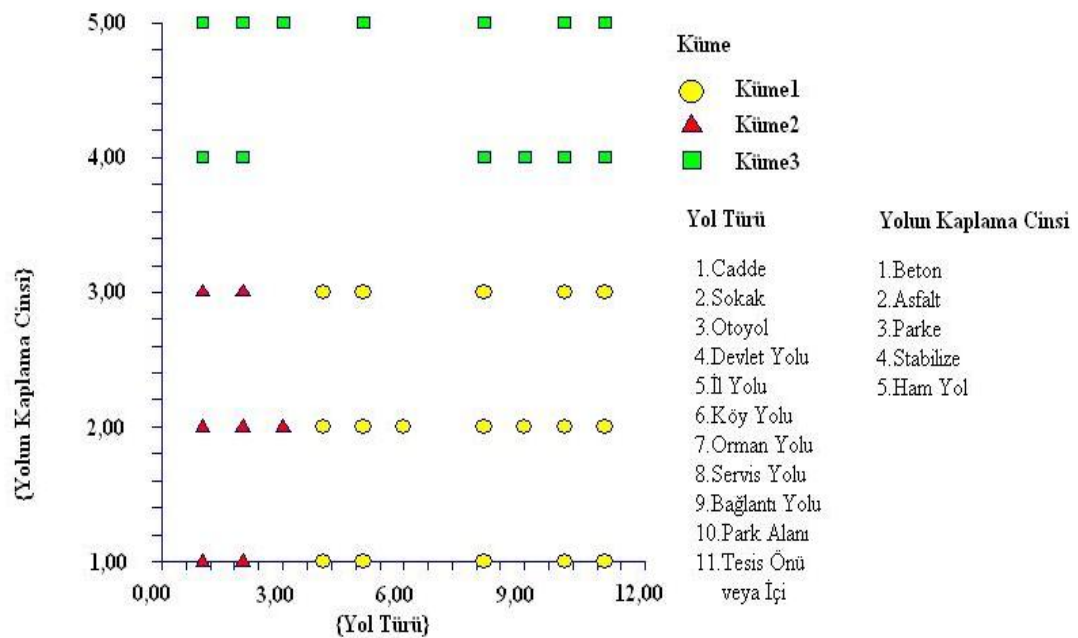
Optimum küme sayısı: 3

Tablo 5.32 Yol türü-yolun kaplama cinsi, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	70,1	
2	2	64,26	
3	2	64,43	
4	3	36,45	
5	3	54,31	
6	3	57,79	

Tablo 5.33 Yol türü- yolun kaplama cinsi, küme merkezleri

Küme Merkez Değişkenleri	Kümeler		
	Küme1	Küme2	Küme3
y	1,9649924	1,9882437	4,2432432
x	5,9619482	1,3425365	3,1756757
Veri Sayısı	657	5614	74



Şekil 5.22 Yol türü-yolun kaplama cinsi, küme grafiği

Küme 1, asfalt kaplamalı il ve köy yollarında, küme 2, asfalt kaplamalı cadde ve sokaklarda, küme 3 ise, stabilize cadde ve sokaklarda kazaların yoğun görüldüğünü göstermektedir. En çok kaza ise asfalt kaplı cadde ve sokaklarda görülmektedir.

5.4.7 Yol bölünme – oluşumuna göre kaza türü

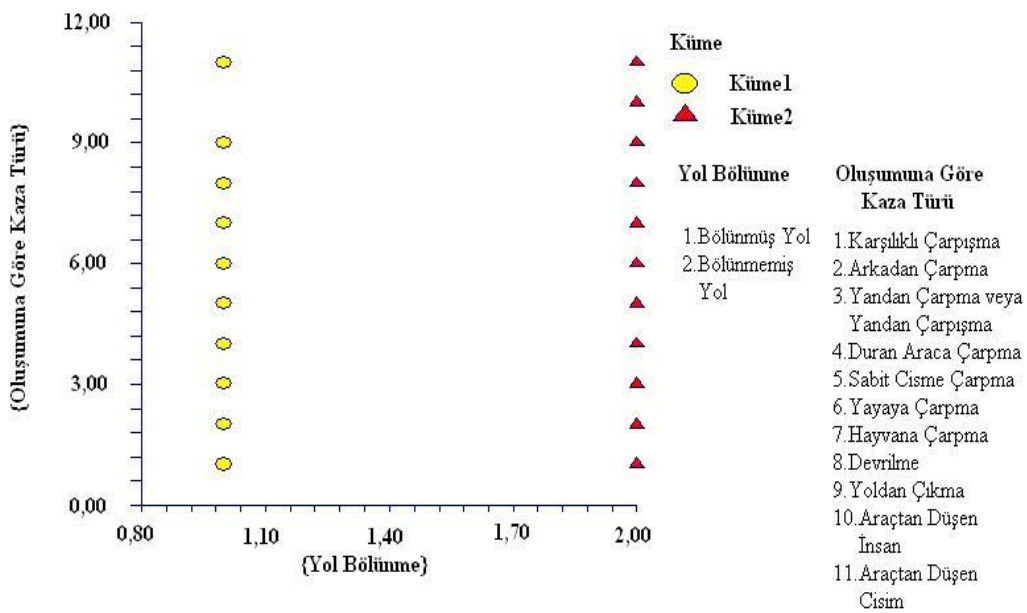
Optimum küme sayısı: 2

Tablo 5.34 Yol bölünme-oluşumuna göre kaza türü, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	69,65	
2	2	49,75	
3	2	49,75	

Tablo 5.35 Yol bölünme-oluşumuna göre kaza türü, küme merkezleri

Küme Merkez Değişkenleri	Kümelere	
	Küme1	Küme2
y	2,9575902	3,1291085
x	1	2
Veri Sayısı	2523	4229



Şekil 5.23 Yol bölünme-oluşumuna göre kaza türü, küme grafiği

Küme 1' de, bölünmüş yollarda daha çok arkadan çarpma veya yandan çarpma, küme 2' de ise, bölünmemiş yollarda yandan çarpma ve duran araca çarpma şeklinde kazaların yoğun olarak meydana geldiği görülmektedir. En çok kaza bölünmemiş yollarda yandan veya arkadan çarpma şeklinde meydana gelmektedir.

5.4.8 Yol bölünme – hava durumu

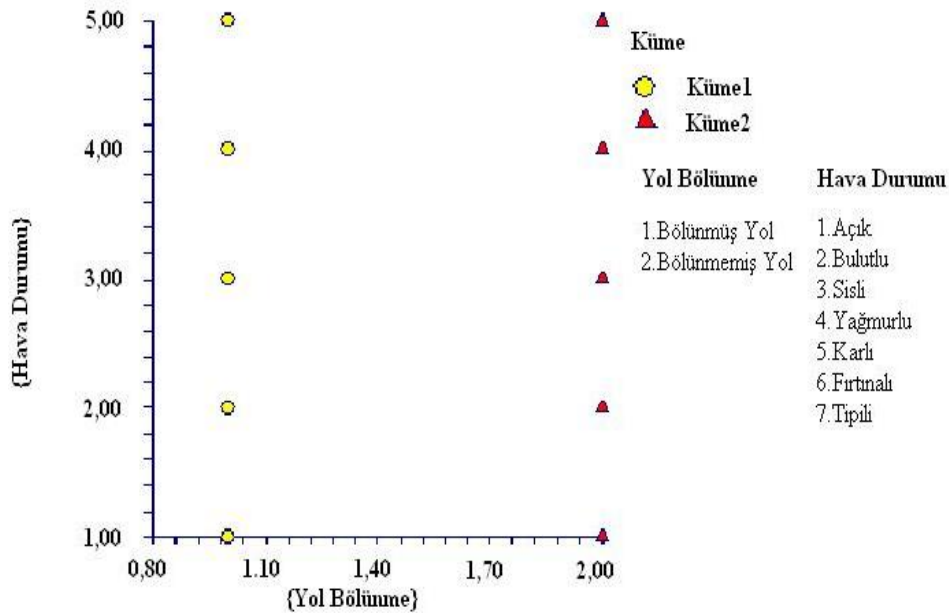
Optimum küme sayısı: 2

Tablo 5.36 Yol bölünme-hava durumu, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	57,81	
2	2	49,94	
3	2	49,94	

Tablo 5.37 Yol bölünme- hava durumu, küme merkezleri

Küme Merkez Değişkenleri	Kümeler	
	Küme1	Küme2
y	1,3406165	1,2852137
x	1	2
Veri Sayısı	2563	4281



Şekil 5.24 Yol bölünme- hava durumu, küme grafiği

Küme 1 ve küme 2' den de anlaşılacağı üzere kazalar bölünmüş ve bölünmemiş yollarda daha çok açık havalarda meydana gelmektedir. Bölünmemiş yollardaki kaza sayısı ise, bölünmüş yollardaki kaza sayısının yaklaşık 1,5 katıdır.

5.4.9 Yol bölünme – gün durumu

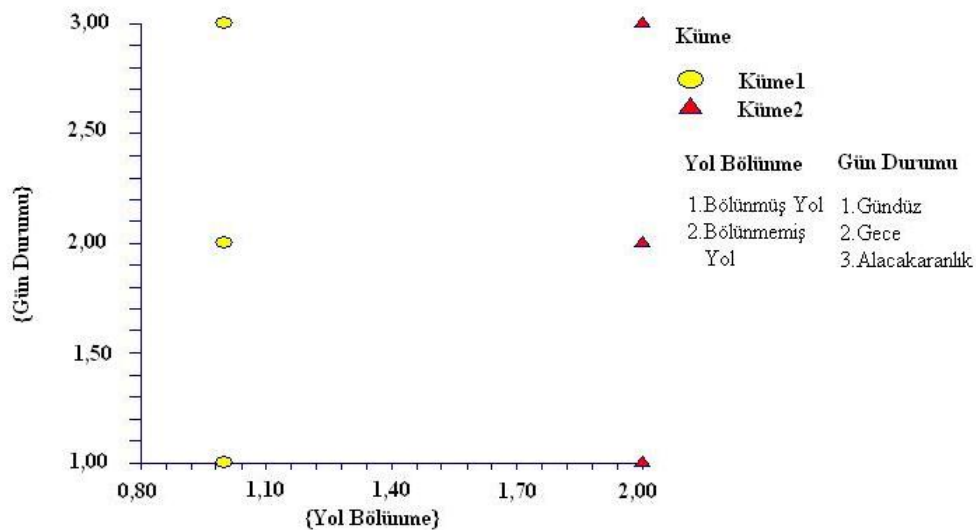
Optimum küme sayısı: 2

Tablo 5.38 Yol bölünme-gün durumu, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	54,44	
2	2	54,44	
3	2	49,94	

Tablo 5.39 Yol bölünme-gün durumu, küme merkezleri

Küme Merkez Değişkenleri	Kümeler	
	Küme1	Küme2
y	1,2724397	1,2388836
x	1	2
Veri Sayısı	2529	4228



Şekil 5.25 Yol bölünme- gün durumu, küme grafiği

Küme merkezlerine bakılacak olursa kazaların büyük çoğunluğunun bölünmüş ve bölünmemiş yollarda gündüzleri meydana geldiği açıkça görülebilir. Bölünmemiş yollardaki kaza sayısı yine bölünmemiş yollardaki kaza sayısının yaklaşık 1,5 katı kadardır.

5.4.10 Yol bölünme – yolun kaplama cinsi

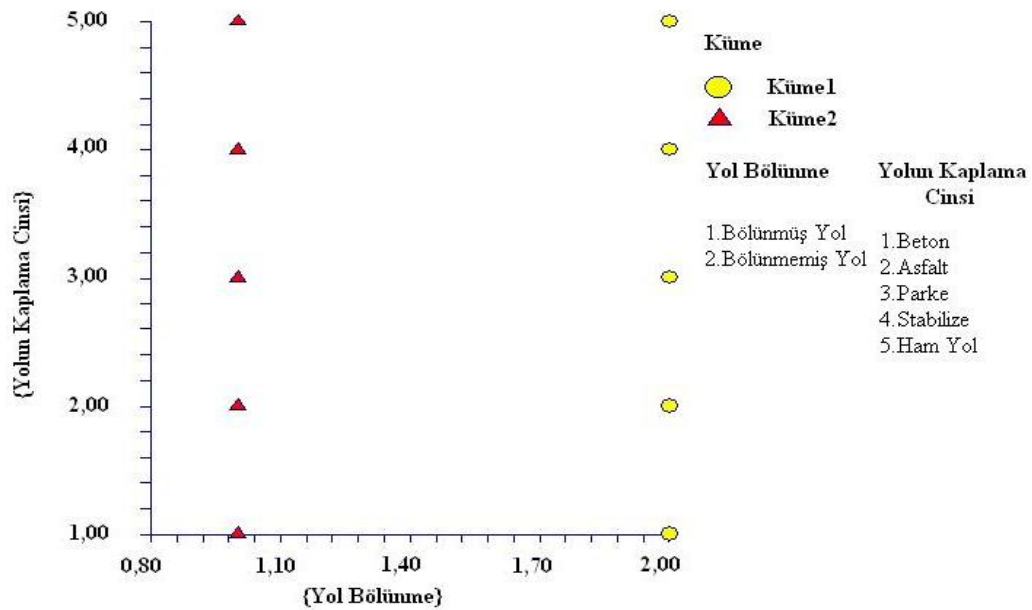
Optimum küme sayısı: 2

Tablo 5.40 Yol bölünme-yolun kaplama cinsi, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	70,94	
2	2	49,95	
3	2	49,95	

Tablo 5.41 Yol bölünme- yolun kaplama cinsi, küme merkezleri

Küme Merkez Değişkenleri	Kümelere	
	Küme1	Küme2
y	2,017924	1,9977679
x	2	1
Veri Sayısı	3738	2240



Şekil 5.26 Yol bölünme- yolun kaplama cinsi, küme grafiği

Kazaların büyük çoğunluğu asfalt kaplamalı bölünmüş ve bölünmemiş yollarda meydana gelmektedir. Bölünmemiş yollardaki kaza sayısının bölünmüş yollara oranla yaklaşık 1,5 kat fazla olduğu burada da açıkça görülmektedir.

5.4.11 Yerleşim durumu – oluşumuna göre kaza türü

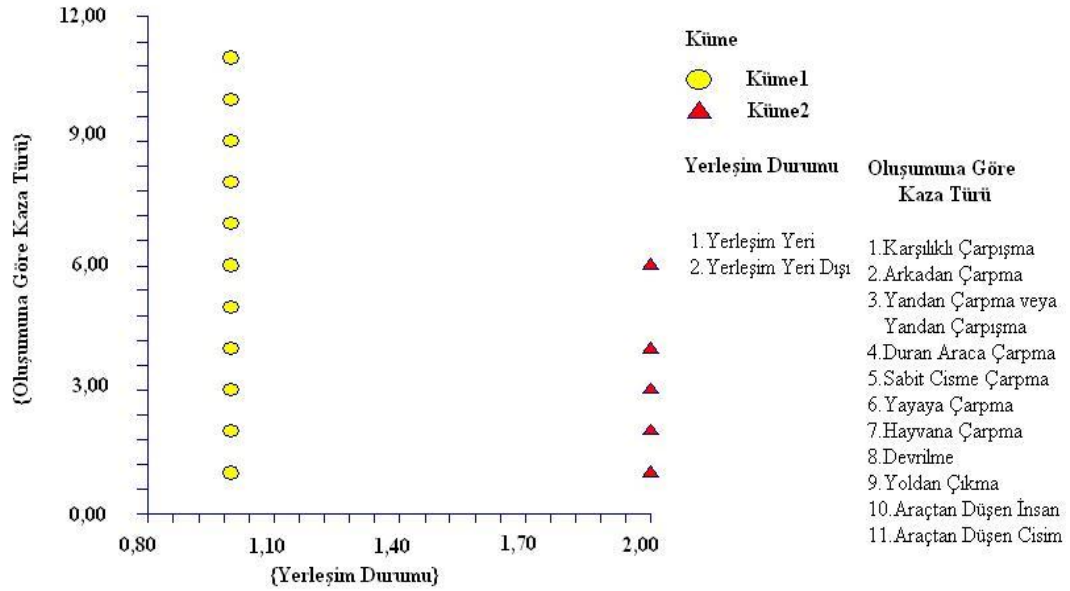
Optimum küme sayısı: 2

Tablo 5.42 Yerleşim durumu-oluşumuna göre kaza türü, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	50	
2	2	80,76	
3	2	69,62	

Tablo5.43 Yerleşim durumu-oluşumuna göre kaza türü, küme merkezleri

Küme Merkez Değişkenleri	Kümeler	
	Küme1	Küme2
y	3,0760532	2,92
x	1	2
Veri Sayısı	7074	25



Şekil 5.27 Yerleşim durumu-oluşumuna göre kaza türü, küme grafiği

Yerleşim yerlerinde kazaların, daha çok yandan çarpma veya çarpışma, yerleşim yerleri dışında ise, daha çok arkadan çarpma, yandan çarpma ve çarpışma şeklinde meydana geldiği Şekil 5.27' den ve küme merkezlerinden görülebilmektedir.

5.4.12 Yerleşim durumu – hava durumu

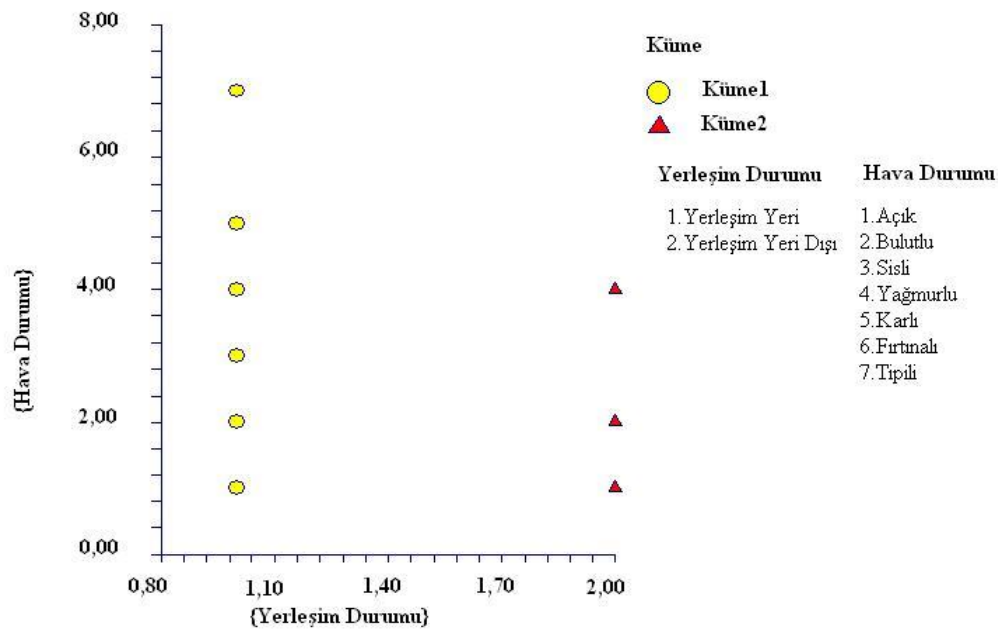
Optimum küme sayısı: 2

Tablo 5.44 Yerleşim durumu-hava durumu, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	50	
2	2	58,6	
3	2	64,56	

Tablo 5.45 Yerleşim durumu-hava durumu, küme merkezleri

Küme Merkez Değişkenleri	Kümeler	
	Küme1	Küme2
y	1,2971805	1,3103448
x	1	2
Veri Sayısı	8796	29



Şekil 5.28 Yerleşim durumu-hava durumu, küme grafiği

Burada yine yerleşim yerleri ve yerleşim yerleri dışında kazaların büyük çoğunluğu açık havada meydana gelmiştir. Yerleşim yerlerinde meydana gelen kaza sayısı ise, yerleşim dışında meydana gelen kaza sayısının çok üstündedir.

5.4.13 Yerleşim durumu – gün durumu

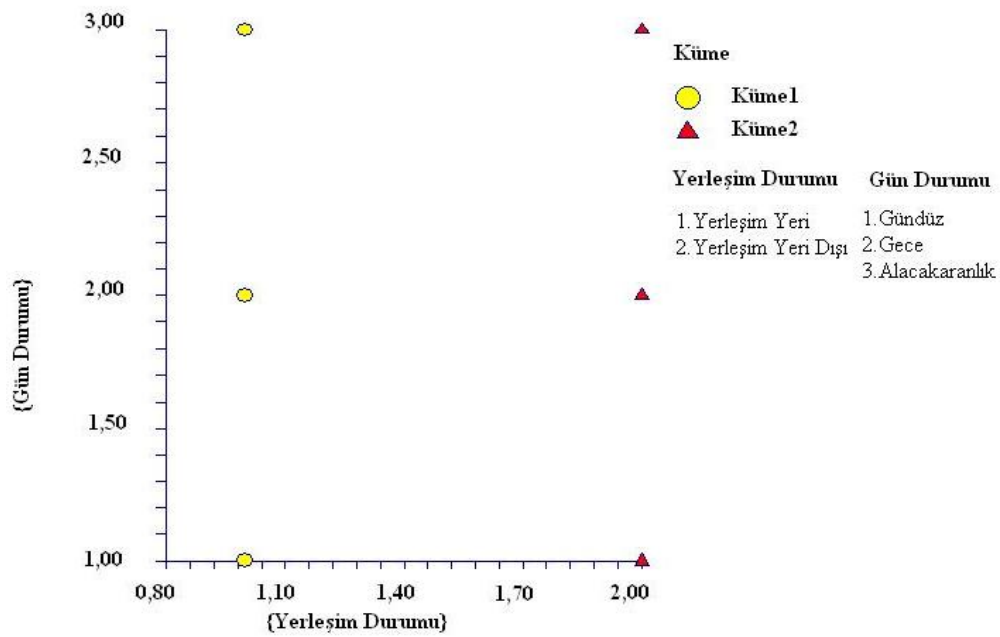
Optimum küme sayısı: 2

Tablo 5.46 Yerleşim durumu-gün durumu, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	49,99	
2	2	54,52	
3	2	54,52	

Tablo 5.47 Yerleşim durumu-gün durumu, küme merkezleri

Küme Merkez Değişkenleri	Kümeler	
	Küme1	Küme2
y	1,2473551	1,3571429
x	1	2
Veri Sayısı	8696	28



Şekil 5.29 Yerleşim durumu-gün durumu, küme grafiği

Yerleşim yerleri ve yerleşim yerleri dışında kazalar büyük oranda gündüzleri meydana gelmiştir. Yine burada, yerleşim yerlerinde meydana gelen kaza sayısı, yerleşim yerleri dışında meydana gelen kaza sayısının çok üstündedir.

5.4.14 Yerleşim durumu – yolda yön

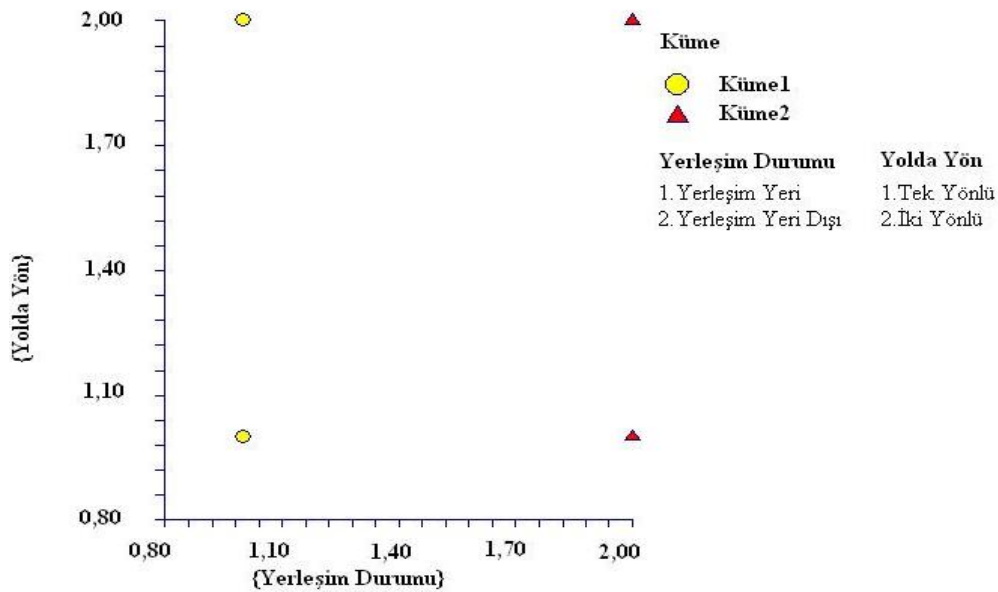
Optimum küme sayısı: 2

Tablo 5.48 Yerleşim durumu-yolda yön, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	49,99	
2	2	49,99	
3	2	49,99	

Tablo 5.49 Yerleşim durumu-yolda yön, küme merkezleri

Küme Merkez Değişkenleri	Kümeler	
	Küme1	Küme2
y	1,6262948	1,72
x	1	2
Veri Sayısı	6275	25



Şekil 5.30 Yerleşim durumu-yolda yön, küme grafiği

Burada, yerleşim yerlerinde ve yerleşim yerleri dışında kazalar genellikle iki yönlü yollarda görülmüştür. En fazla kaza ise yerleşim yerlerindeki iki yönlü yollarda meydana gelmiştir.

5.4.15 Yerleşim durumu – yolun kaplama cinsi

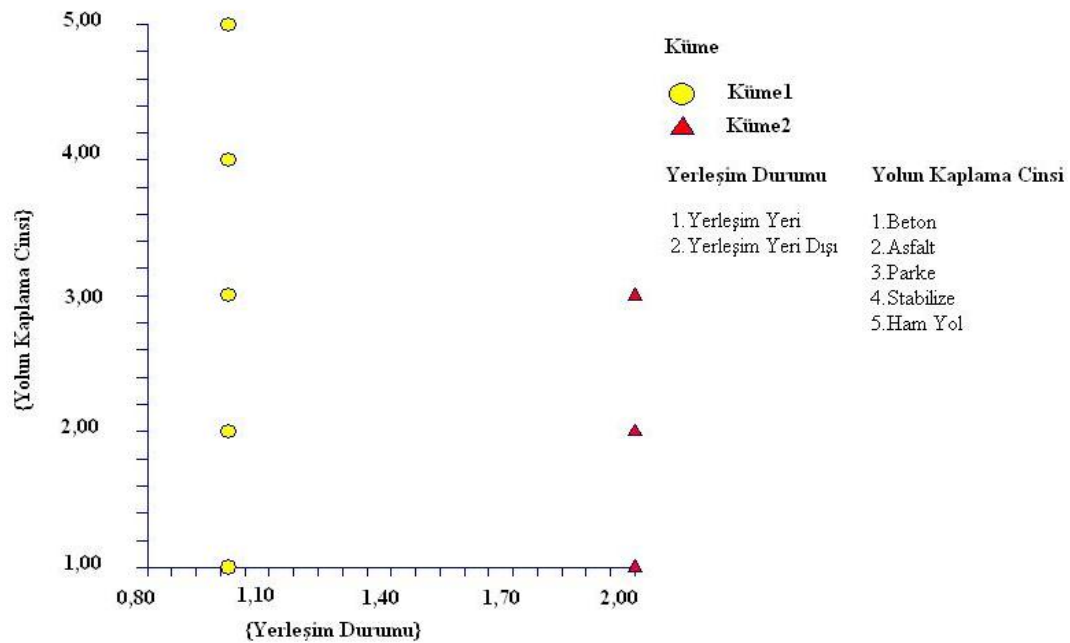
Optimum küme sayısı: 2

Tablo 5.50 Yerleşim durumu-yolun kaplama cinsi, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	50	
2	2	85,71	
3	2	50	

Tablo 5.51 Yerleşim durumu- yolun kaplama cinsi, küme merkezleri

Küme Merkez Değişkenleri	Kümeler	
	Küme1	Küme2
y	2,0108666	2
x	1	2
Veri Sayısı	7362	28



Şekil 5.31 Yerleşim durumu-yolun kaplama cinsi, küme grafiği

Bölünmüş ve bölünmemiş yollardaki kazaların yine genellikle asfalt kaplamalı yollarda meydana geldiği açıkça görülmektedir.

5.4.16 Hava durumu – oluşumuna göre kaza türü

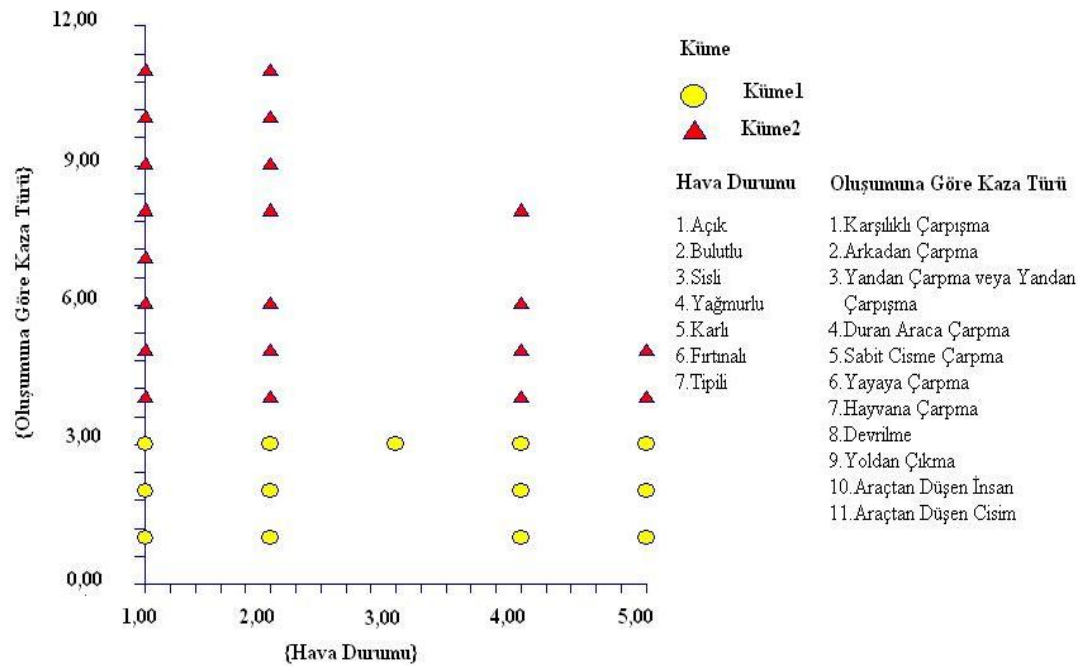
Optimum küme sayısı: 2

Tablo 5.52 Hava durumu-oluşumuna göre kaza türü, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	69,57	
2	2	75,41	
3	2	69,39	

Tablo 5.53 Hava durumu-oluşumuna göre kaza türü, küme merkezleri

Küme Merkez Değişkenleri	Kümeler	
	Küme1	Küme2
y	2,6290432	4,9399718
x	1,3122234	1,2831921
Veri Sayısı	5874	1416



Şekil 5.32 Hava durumu-oluşumuna göre kaza türü, küme grafiği

Burada, genellikle kazaların açık havalarda arkadan çarpma, yandan çarpma veya çarpışma başta olmak üzere, duran araca çarpma ve sabit cisme çarpma şeklinde meydana geldiği görülmektedir.

5.4.17 Hava durumu – yolun kaplama cinsi

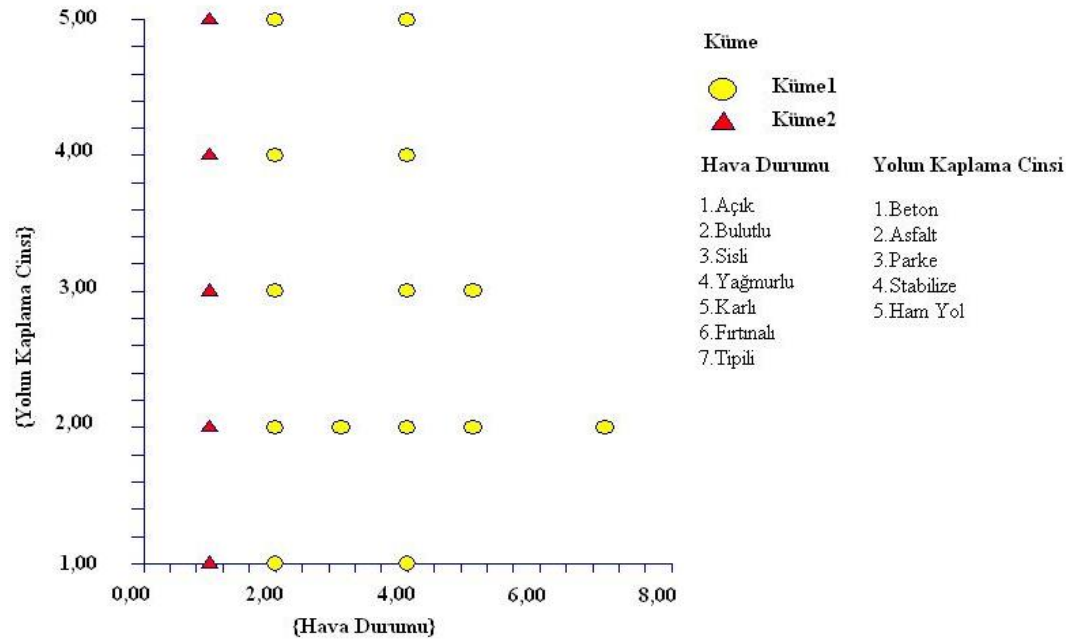
Optimum küme sayısı: 2

Tablo 5.54 Hava durumu-yolun kaplama cinsi, iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	70,92	
2	2	68,79	
3	2	64,55	

Tablo 5.55 Hava durumu- yolun kaplama cinsi, küme merkezleri

Küme Merkez Değişkenleri	Kümeler	
	Küme1	Küme2
y	2,0072934	2,0118521
x	2,6961102	1
Veri Sayısı	1234	6328



Şekil 5.33 Hava durumu-yolun kaplama cinsi, küme grafiği

Burada açık, bulutlu ve yağmurlu havalarda kazaların genelde asfalt kaplamalı yollarda meydana geldiği görülmektedir.

5.5 Alt Kümeleme Analizleri

Alt kümeleme analizinde, kümeleme parametrelerinden sadece oluşumuna göre kaza türünün alt dalları olan;

- Karşılıklı çarpışma
- Arkadan Çarpma
- Yandan çarpma veya yandan çarpışma
- Duran araca çarpma
- Sabit cisme çarpma
- Yayaya çarpma
- Hayvana çarpma
- Devrilme
- Yoldan çıkma
- Araçtan düşen insan
- Araçtan düşen cisim

parametreleri ele alınmış olup, bu parametrelere denk gelen koordinat verileri seçilerek, k-ortalamlar yöntemi ile kümeleme analizi yapılmıştır. Bu analizde hangi bölgelerde hangi tür kazaların daha sık görüldüğünün belirlenmesi amaçlanmıştır.

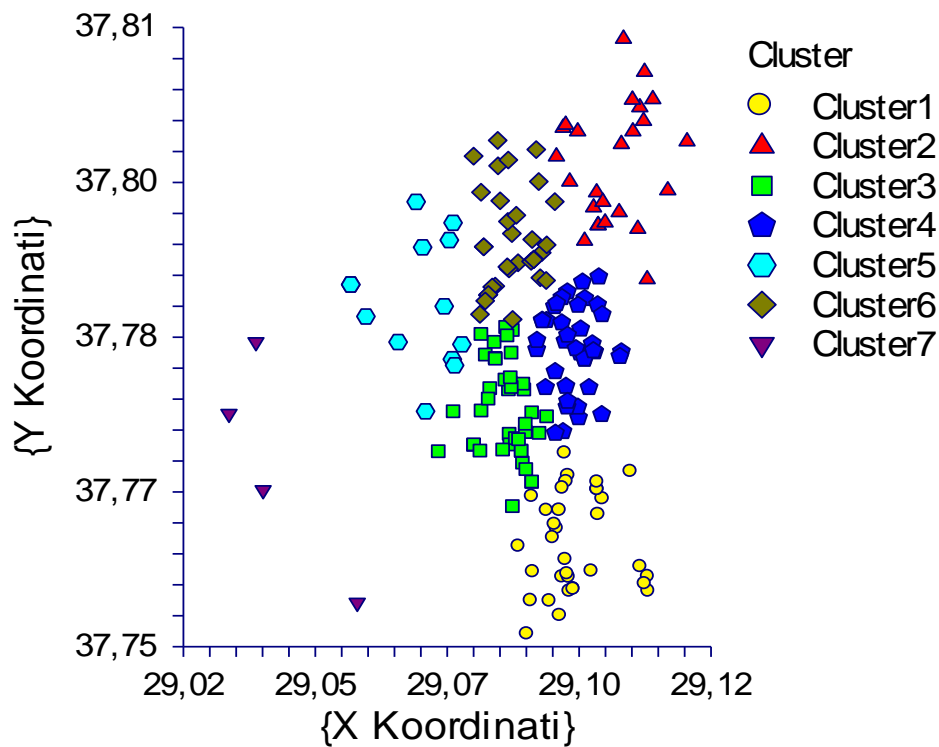
5.5.1 Karşılıklı çarpışma

Tablo 5.56 Koordinat verisi (karşılıklı çarpışma), iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	72,28	
2	2	65,87	
3	2	65,87	
4	3	43,91	
5	3	41,9	
6	3	41,9	
7	4	31,04	
8	4	31,04	
9	4	31,04	
10	5	24,19	
11	5	24,19	
12	5	24,35	
13	6	18	
14	6	18	
15	6	18	
16	7	15,76	
17	7	16,33	
18	7	16	

Tablo 5.57 Koordinat verisi (karşılıklı çarpışma), küme merkezleri

Küme No	Değişkenler		Veri Sayısı
	x	y	
1	29,09462	37,75987	34
2	29,1017	37,79676	25
3	29,08113	37,77294	41
4	29,09398	37,77894	39
5	29,06462	37,78339	13
6	29,08235	37,78934	33
7	29,0376	37,76787	4

**Şekil 5.34** Koordinat verisi (karşılıklı çarpışma), küme grafiği

Bu analiz sonucunda, küme merkezlerine göre elde edilen kara noktalar ise Tablo 5.58’ de verilmiştir.

Tablo 5.58’ den görüldüğü üzere karşılıklı çarpışma türü kazalarda ilk kara noktayı Lise, Çaybaşı ve Kıbrıs Şehitler Caddeler ‘inin kesişim noktası olan lise kavşağı meydana getirmektedir. Bu tür kazalar daha çok bölünmemiş yollarda hatalı sollamalar ve hız ihlali sonucu meydana gelmektedir.

Tablo 5.58 Karşılıklı çarpışmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
Lise ve Çaybaşı Caddesi	3	41
Atatürk Caddesi	4	39
Demokrasi Meydanı	1	34
Örnek Caddesi	6	33
Sevindik	2	25
29 Ekim Bulvarı	5	13
Yeni Adliye	7	4

5.5.2 Arkadan çarpma

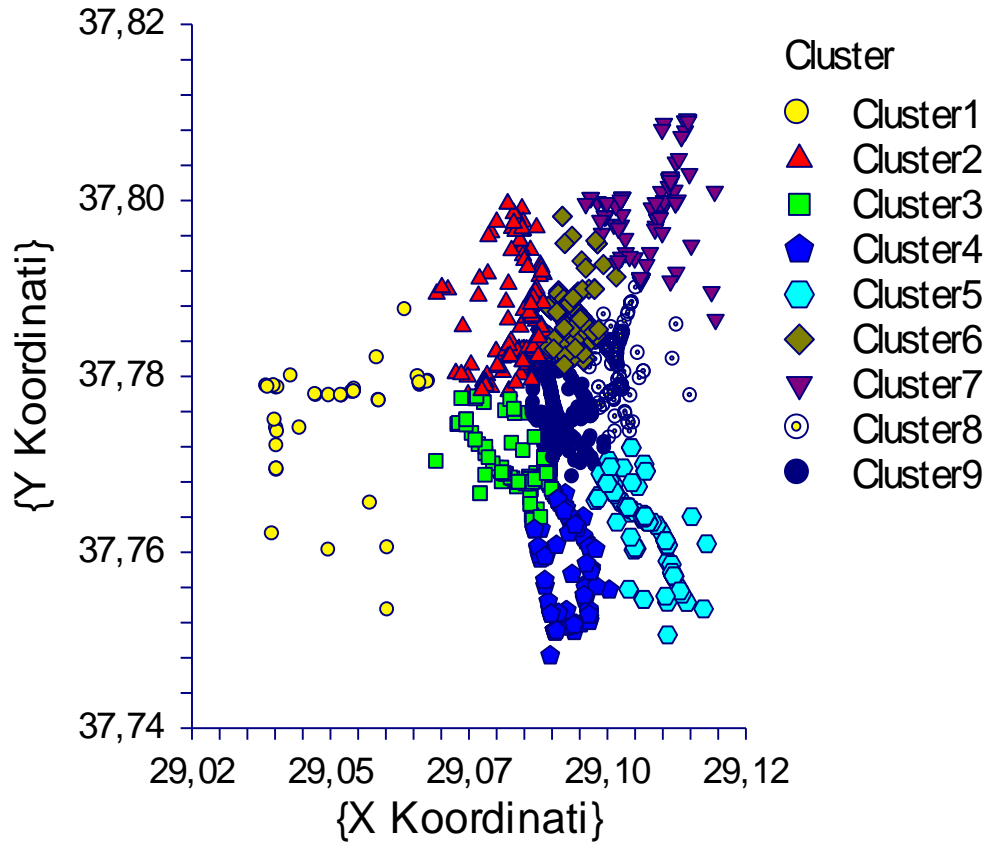
Optimum küme sayısı: 9

Tablo 5.59 Koordinat verisi (arkadan çarpma), iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	77,2	
2	2	66,25	
3	2	77,2	
4	3	47,35	
5	3	54,88	
6	3	47,26	
7	4	34,83	
8	4	40,83	
9	4	36,33	
10	5	29,19	
11	5	27,9	
12	5	27,93	
13	6	23,8	
14	6	20,04	
15	6	20,04	
16	7	17,17	
17	7	16,96	
18	7	19,42	
19	8	14,77	
20	8	14,74	
21	8	14,64	
22	9	12,84	
23	9	12,86	
24	9	12,66	

Tablo 5.60 Koordinat verisi (arkadan çarpma), küme merkezleri

Küme No	Değişkenler		Veri Sayısı
	x	y	
1	29,04738	37,77528	43
2	29,07926	37,7865	235
3	29,07972	37,76959	153
4	29,08768	37,75924	135
5	29,1007	37,7636	127
6	29,08879	37,78641	353
7	29,10332	37,79929	116
8	29,09718	37,78127	294
9	29,08698	37,77529	349



Şekil 5.35 Koordinat verisi (arkadan çarpma), küme grafiği

Analiz sonucu elde edilen kara noktalar Tablo 5.61' deki gibidir;

Tablo 5.61 Arkadan çarpmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
İzmir Bulvarı (İstasyon Kavşağı)	6	353
Çınar Meydanı	9	349
Ulusal Egemenlik Meydanı	8	294
Devlet Hastanesi	2	235
Lise	3	153
Kıbrıs Şehitleri Caddesi	4	135
Fatih Kavşağı	5	127
Sevindik	7	116
Yeni Adliye	1	43

En çok arkadan çarpma İstasyon Kavşağı' nda meydana gelmiştir. Bu tür kazalar genel olarak sürücülerin sinyalizasyonlu kavşaklarda, ışığın sarıdan kırmızıya geçişi sırasında ani olarak durmasından dolayı meydana gelmektedir.

5.5.3 Yandan çarpma veya yandan çarpışma

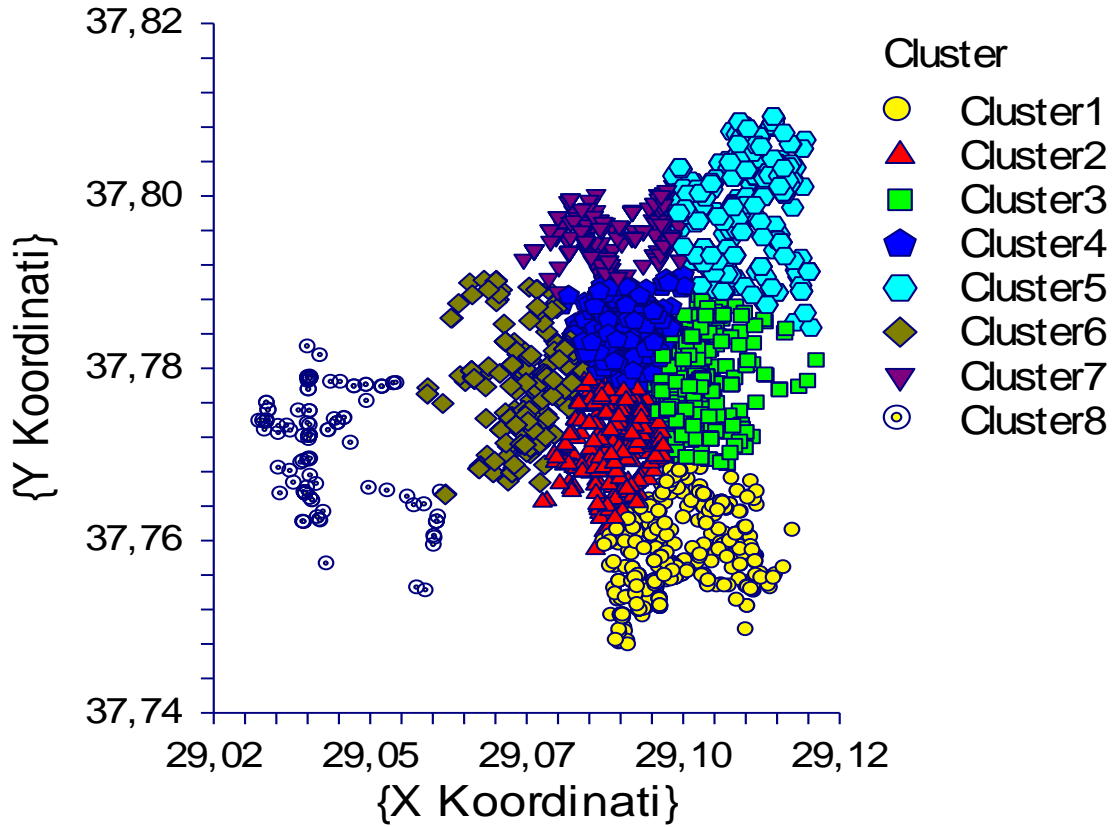
Optimum küme sayısı: 8

Tablo 5.62 Koordinat verisi (yandan çarpma), iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	73,71	
2	2	64,24	
3	2	64,24	
4	3	42,53	
5	3	42,39	
6	3	42,53	
7	4	29,53	
8	4	29,53	
9	4	29,53	
10	5	22,97	
11	5	23,89	
12	5	22,98	
13	6	17,32	
14	6	17,31	
15	6	17,34	
16	7	14,42	
17	7	15,5	
18	7	14,42	
19	8	12,69	
20	8	12,81	
21	8	12,69	

Tablo 5.63 Koordinat verisi (yandan çarpma), küme merkezleri

Küme No	Değişkenler		Veri Sayısı
	x	y	
1	29,094878	37,759085	562
2	29,083885	37,771437	551
3	29,097368	37,778700	593
4	29,086410	37,784919	964
5	29,104290	37,798428	453
6	29,070148	37,778207	364
7	29,084413	37,795453	289
8	29,038491	37,770879	122



Şekil 5.36 Koordinat verisi (yandan çarpma), küme grafiği

Analiz sonucu elde edilen kara noktalar ise Tablo 5.64'deki gibidir;

Tablo 5.64 Yandan çarpmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
İzmir Bulvarı (İstasyon Kavşağı)	4	964
Deliktaş Kavşağı	3	593
Fatih Kavşağı	1	562
Lise	2	551
Sevindik	5	453
Özay Gönüm Meydanı	6	364
İzmir Bulvarı (Zafer Katrancı Kavşağı)	7	289
Yeni Adliye	8	122

Yandan çarpma veya çarpışma türü kaza yine en çok İstasyon Kavşağı'nda meydana gelmiştir. Bu tür kazalar sürücülerin ışıklarda dikkatsiz davranıp, henüz yeşil yanmadan hareket etmesinden veya kırmızı yandığı halde geçmek istemesinden dolayı meydana gelmektedir.

5.5.4 Duran araca çarpma

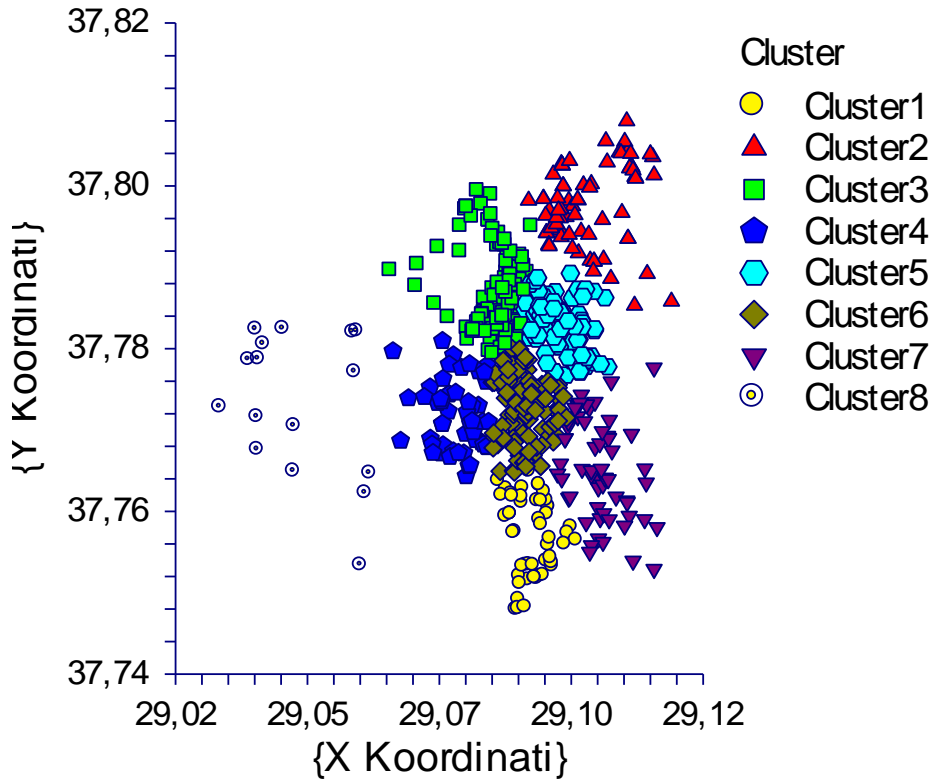
Optimum küme sayısı: 8

Tablo 5.65 Koordinat verisi (duran araca çarpma), iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	75,67	
2	2	66,99	
3	2	66,99	
4	3	46,71	
5	3	47,09	
6	3	47,12	
7	4	34,59	
8	4	35,53	
9	4	34,59	
10	5	28,61	
11	5	27,55	
12	5	27,7	
13	6	25,6	
14	6	21,83	
15	6	21,72	
16	7	18,54	
17	7	18,67	
18	7	18,84	
19	8	15,45	
20	8	15,58	
21	8	15,44	

Tablo 5.66 Koordinat verisi (duran araca çarpma), küme merkezleri

Küme No	Değişkenler		Veri Sayısı
	x	y	
1	29,08788	37,75658	54
2	29,0983	37,79742	71
3	29,08034	37,7871	120
4	29,07331	37,7719	52
5	29,09207	37,78315	119
6	29,086	37,77318	159
7	29,10044	37,76527	62
8	29,04371	37,77309	16



Şekil 5.37 Koordinat verisi (duran araca çarpma), küme grafiği

Analiz sonucu elde edilen kara noktalar Tablo 5.67' deki gibidir;

Tablo 5.67 Duran araca çarpmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
Çınar Meydanı	6	159
Örnek Caddesi	3	120
İzmir Bulvarı (İstasyon Kavşağı)	5	119
25. Sokak	2	71
Fatih Kavşağı	7	62
Demokrasi Meydanı	1	54
Fatih Caddesi ve Dumlupınar Caddesi	4	52
Yeni Adliye	8	16

Duran araca çarpma bakımında ilk kara nokta, merkezde bulunan ve yoğun trafiğin olduğu Çınar Meydanı' dır. Bu tür kazalara genel olarak sürücülerin dikkatsizliği ve kontrolsüzce bırakılmış araçlar sebebiyet vermektedir.

5.5.5 Sabit cisme çarpma

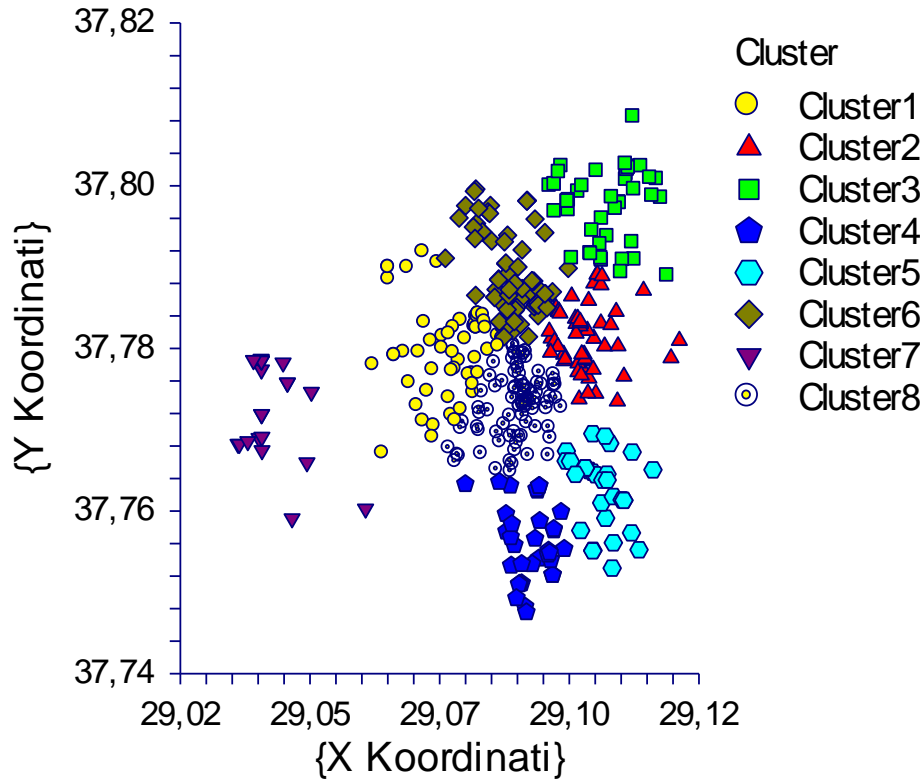
Optimum küme sayısı: 8

Tablo 5.68 Koordinat verisi (sabit cisme çarpma), iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	71,88	
2	2	65,13	
3	2	65,13	
4	3	40,74	
5	3	40,79	
6	3	40,79	
7	4	28,7	
8	4	29,34	
9	4	29,3	
10	5	22,07	
11	5	22,07	
12	5	22,61	
13	6	17,3	
14	6	17,18	
15	6	17,22	
16	7	15,13	
17	7	15,1	
18	7	15,07	
19	8	13,64	
20	8	14,12	
21	8	12,59	

Tablo 5.69 Koordinat verisi (sabit cisme çarpma), küme merkezleri

Küme No	Değişkenler		Veri Sayısı
	x	y	
1	29,072714	37,780000	64
2	29,098273	37,781389	57
3	29,101769	37,797641	47
4	29,087410	37,756219	34
5	29,101332	37,762790	35
6	29,084080	37,788997	77
7	29,037784	37,772079	21
8	29,085349	37,773066	92



Şekil 5.38 Koordinat verisi (sabit cisme çarpma), küme grafiği

Analiz sonucu elde edilen kara noktalar Tablo 5.70' deki gibidir;

Tablo 5.70 Sabit cisme çarpmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
Çınar Meydanı	8	92
İzmir Bulvarı (İstasyon Kavşağı)	6	77
Özay Gönlüm Meydanı	1	64
Deliktaş Kavşağı	2	57
Sevindik	3	47
Kıbrıs Şehitleri Caddesi	5	35
Demokrasi Meydanı	4	34
Yeni Adliye	7	21

Sabit cisme çarpma bakımından en çok kaza yine Çınar Meydanı' nda meydana gelmiştir. Bu tür kazalar da genel olarak sürücü dikkatsizliği ve yüksek hız sonucu meydana gelmektedir.

5.5.6 Yayaya çarpma

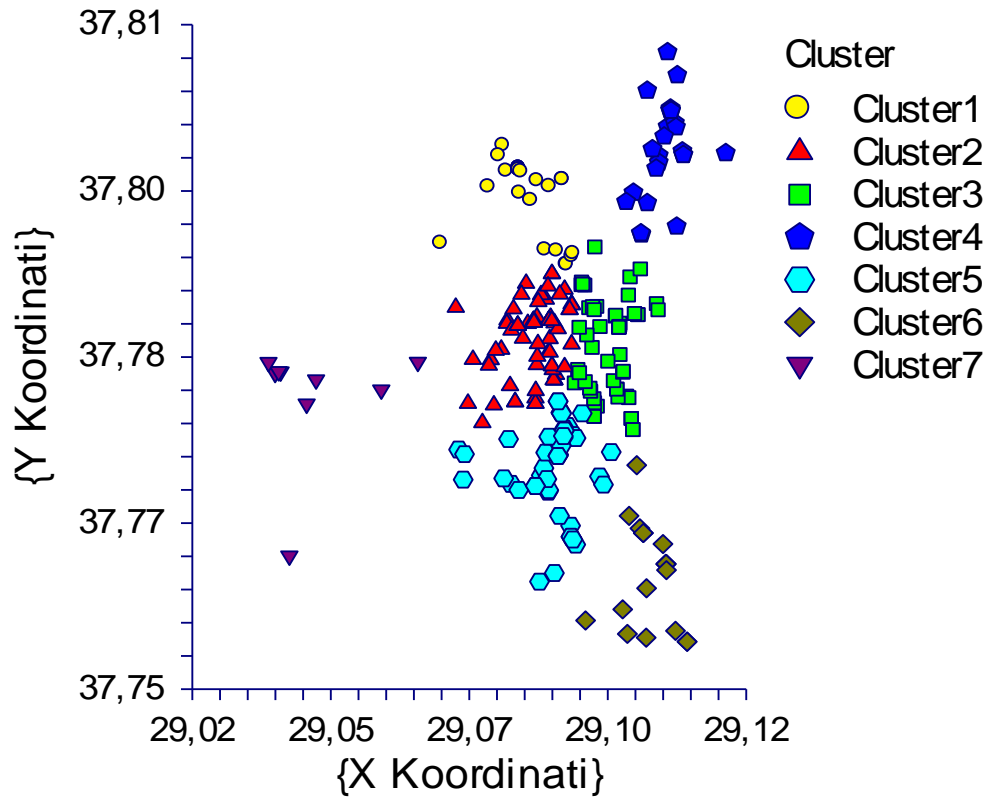
Optimum küme sayısı: 7

Tablo 5.71 Koordinat verisi (yayaya çarpma), iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	76,67	
2	2	67,34	
3	2	67,34	
4	3	58,15	
5	3	60,36	
6	3	47,62	
7	4	32	
8	4	31,86	
9	4	33	
10	5	25,13	
11	5	25,13	
12	5	25,4	
13	6	16,62	
14	6	16,79	
15	6	17,07	
16	7	13,27	
17	7	13,55	
18	7	13,55	

Tablo 5.72 Koordinat verisi (yayaya çarpma), küme merkezleri

Küme No	Değişkenler		Veri Sayısı
	x	y	
1	29,081223	37,794233	24
2	29,081691	37,781509	68
3	29,095416	37,780968	61
4	29,105368	37,798964	26
5	29,084853	37,770090	46
6	29,102763	37,760186	19
7	29,041718	37,776397	9



Şekil 5.39 Koordinat verisi (yayaya çarpma), küme grafiği

Analiz sonucu elde edilen kara noktalar Tablo 5.73' deki gibidir;

Tablo 5.73 Yayaya çarpmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
Devlet Hastanesi	2	68
Ulusal Egemenlik Meydanı	3	61
Çınar Meydanı	5	46
Sevindik	4	26
İzmir Bulvarı (Zafer Katrancı Kavşağı)	1	24
Fatih Kavşağı	6	19
Yeni Adliye	7	9

Görüldüğü üzere en çok yayaya çarpma, yaya trafiğinin en yoğun olduğu hastane civarında meydana gelmiştir. Yayaya çarpma kazaları daha çok yayaların kusurlu davranışları, sürücülerin dikkatsizlikleri ve hız kurallarına uymaması sonucu meydana gelmektedir.

5.5.7 Hayvana çarpma

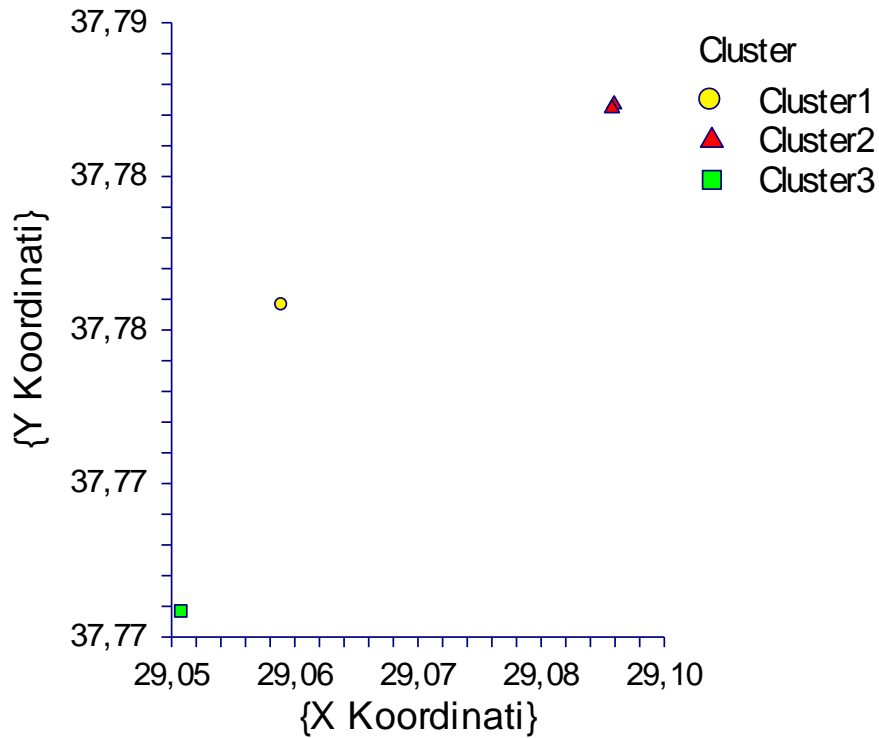
Optimum küme sayısı: 3

Tablo 5.74 Koordinat verisi (hayvana çarpma), iterasyon

İterasyon Bölümü			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	15,34	
2	2	15,34	
3	2	15,34	
4	3	0	
5	3	0	
6	3	0	

Tablo 5.75 Koordinat verisi (hayvana çarpma), küme merkezleri

Küme Merkez Değişkenleri	Kümelere		
	Küme1	Küme2	Küme3
Y	37,778449	37,786594	37,766058
X	29,060197	29,090319	29,050835
Hesaplanan	1	2	1



Şekil 5.40 Koordinat verisi (hayvana çarpma), küme grafiği

Analiz sonucu Tablo 5.76' daki gibidir;

Tablo 5.76 Hayvana çarpmaya ait konum verisinin kümeleme analizi sonucu

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
İzmir Bulvarı (İstasyon Kavşağı)	2	2
29 Ekim Bulvarı	1	1
Meclis Caddesi	3	1

Tablo 5.76' dan görüldüğü üzere hayvana çarpma türü kaza sayısı oldukça azdır. Üç yıl boyunca toplamda 4 noktada bu tür kazaya rastlanmıştır.

5.5.8 Devrilme

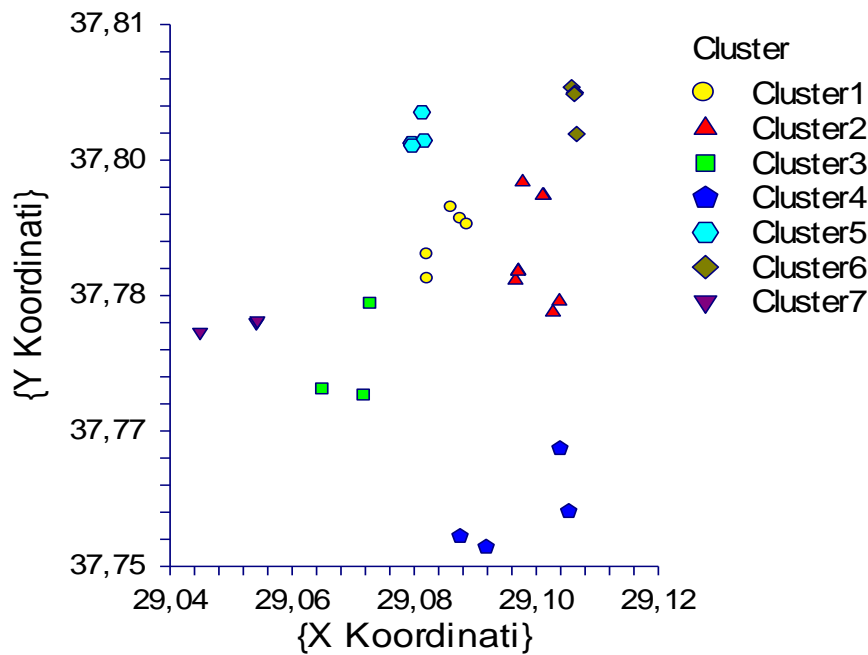
Optimum küme sayısı: 7

Tablo 5.77 Koordinat verisi (devrilme), iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	55,5	
2	2	55,5	
3	2	55,5	
4	3	30,08	
5	3	46,54	
6	3	30,08	
7	4	23,11	
8	4	21,11	
9	4	18,62	
10	5	12,14	
11	5	12,14	
12	5	12,81	
13	6	9,13	
14	6	8,9	
15	6	9,11	
16	7	5,98	
17	7	6	
18	7	9,05	

Tablo 5.78 Koordinat verisi (devrilme), küme merkezleri

Küme No	Değişkenler		Veri Sayısı
	x	y	
1	29,085577	37,786346	5
2	29,099652	37,784913	8
3	29,069721	37,772639	3
4	29,097122	37,756170	4
5	29,080321	37,797548	5
6	29,106248	37,801418	4
7	29,051113	37,776710	3

**Şekil 5.41** Koordinat verisi (devrilme), küme grafiği

Analiz sonucu elde edilen kara noktalar Tablo 5.79' daki gibidir;

Tablo 5.79 Devrilmeye ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
Ulusal Egemenlik Meydanı	2	8
İzmir Bulvarı (İstasyon Kavşağı)	1	5
İzmir Bulvarı (Zafer Katrancı Kavşağı)	5	5
Demokrasi Meydanı	4	4
Sevindik	6	4
Fatih Caddesi	3	3
Yeni Adliye	7	3

En çok devrilme türü kaza İzmir, Ankara ve Antalya yollarını birbirine bağlayan ve trafiğin çok yoğun olduğu Ulusal Egemenlik Meydanı'nda (Üçgen) meydana gelmiştir. Bu tür kazalar daha çok iklim koşullarının yollara etkisi, tekerlek patlaması, dengesiz yükleme ve viraja hızlı girme sonucu meydana gelmektedir.

5.5.9 Yoldan çıkma

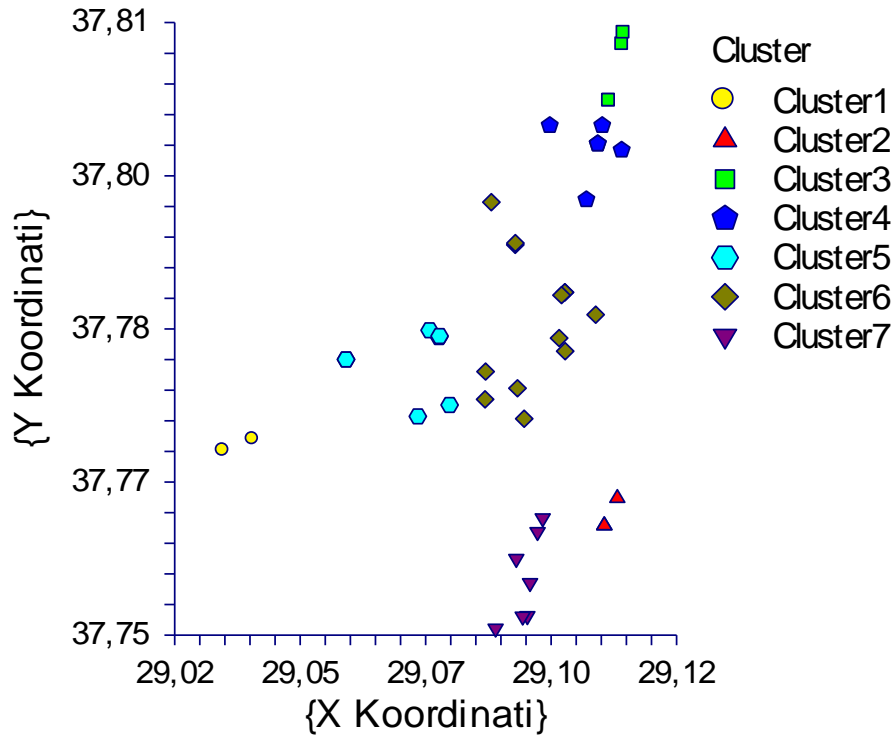
Optimum küme sayısı: 7

Tablo 5.80 Koordinat verisi (yoldan çıkma), iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	52,11	
2	2	52,11	
3	2	52,11	
4	3	30,05	
5	3	27,81	
6	3	27,81	
7	4	16,65	
8	4	16,65	
9	4	18,69	
10	5	10,79	
11	5	16,93	
12	5	10,79	
13	6	8,79	
14	6	13,61	
15	6	8,89	
16	7	8,08	
17	7	8,08	
18	7	7,96	

Tablo 5.81 Koordinat verisi (yoldan çıkma), küme merkezleri

Küme No	Değişkenler		Veri Sayısı
	x	y	
1	29,032841	37,768512	2
2	29,106462	37,761630	3
3	29,107720	37,805493	4
4	29,103262	37,797713	6
5	29,066812	37,776614	7
6	29,091649	37,780922	13
7	29,089720	37,755495	7



Şekil 5.42 Koordinat verisi (yoldan çıkma), küme grafiği

Analiz sonucu elde edilen kara noktalar Tablo 5.82' deki gibidir;

Tablo 5.82 Yoldan çıkmaya ait konum verisinin kümeleme analizi sonucu belirlenen kara noktalar

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
Ulusal Egemenlik Meydanı	6	13
Özay Gönülüm Meydanı	5	7
Demokrasi Meydanı	7	7
Dokuzkavaklar Alt Geçidi	4	6
Sevindik	3	4
Fevzi Çakmak Bulvarı	2	3
Yeni Adliye	1	2

Ulusal Egemenlik Meydanı yoldan çıkma kaza türü bakımından yine ilk sırada yer almıştır. Bu tür kazalar da genel olarak yüksek hız ve dikkatsizlik sonucu meydana gelmektedir.

5.5.10 Araçtan düşen insan

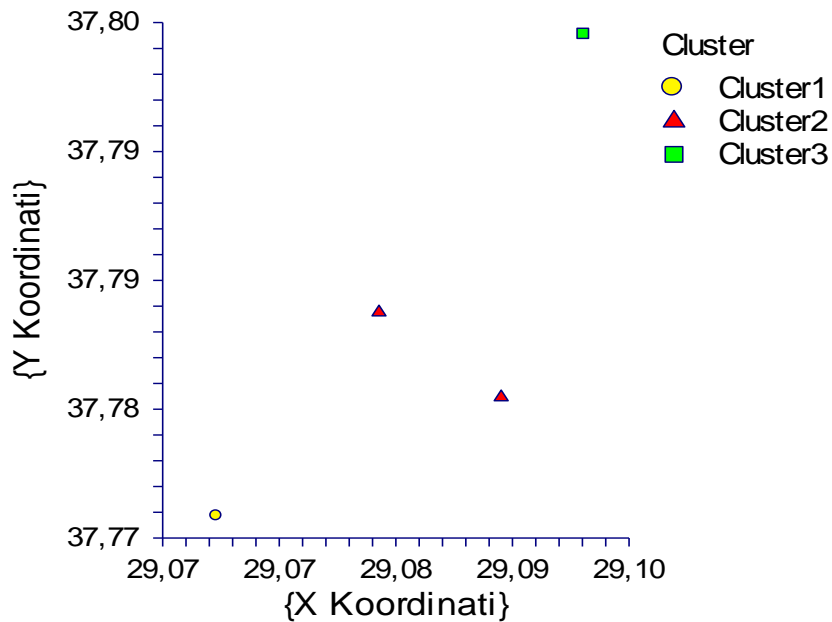
Optimum küme sayısı: 3

Tablo 5.83 Koordinat verisi (araçtan düşen insan), iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	35,02	
2	2	35,02	
3	2	35,02	
4	3	6,36	
5	3	6,36	
6	3	6,36	

Tablo 5.84 Koordinat verisi (araçtan düşen insan), küme merkezleri

Küme Merkez Değişkenleri	Kümeler		
	Küme1	Küme2	Küme3
Y	37,771213	37,780662	37,799373
X	29,069145	29,085816	29,096531
Hesaplanan	1	2	1



Şekil 5.43 Koordinat verisi (araçtan düşen insan), küme grafiği

Analiz sonucu Tablo 5.85' deki gibidir;

Tablo 5.85 Araçtan düşen insana ait konum verisinin kümeleme analizi sonucu

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
Vilayet Önü Kavşak	2	2
Yeşilköy Caddesi	1	1
25. Sokak	3	1

5.5.11 Araçtan düşen cisim

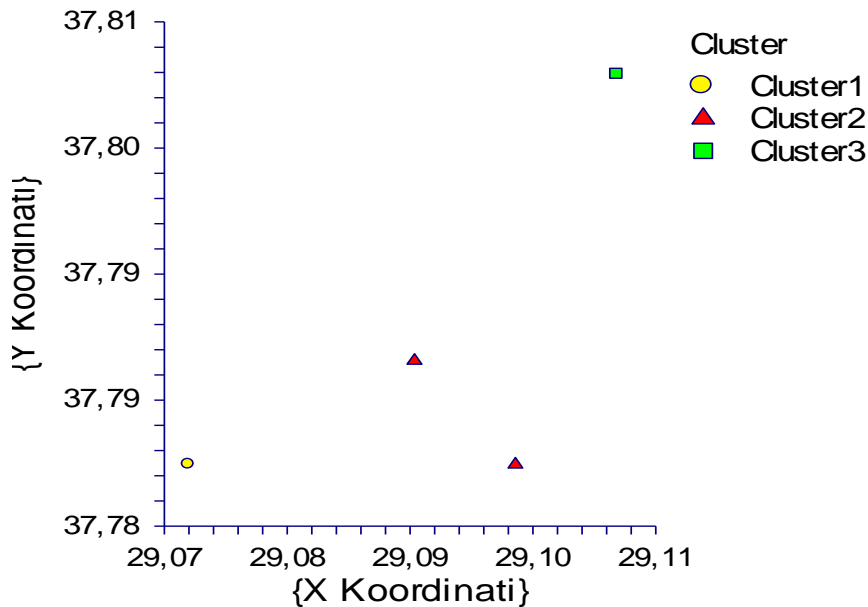
Optimum küme sayısı: 3

Tablo 5.86 Koordinat verisi (araçtan düşen cisim), iterasyon

İterasyon			
İterasyon No	Küme No	Değişim Yüzdesi	Yüzdenin Çizgisel Grafiği
1	2	31,4	
2	2	50,13	
3	2	31,4	
4	3	5,18	
5	3	5,18	
6	3	5,18	

Tablo 5.87 Koordinat verisi (araçtan düşen cisim), küme merkezleri

Küme Merkez Değişkenleri	Kümeler		
	Küme1	Küme2	Küme3
Y	37,783015	37,785628	37,80245
X	29,067319	29,092527	29,106341
Hesaplanan	1	2	1

**Şekil 5.44** Koordinat verisi (araçtan düşen cisim), küme grafiği

Analiz sonucu Tablo 5.88' deki gibidir;

Tablo 5.88 Araçtan düşen cisme ait konum verisinin kümeleme analizi sonucu

Kavşak veya Cadde (Kara Nokta)	Küme No	Kaza Sayısı
İzmir Bulvarı (İstasyon Kavşağı)	2	2
201. Sokak	1	1
Sevindik Alt Geçidi	3	1

6. SONUÇLAR VE ÖNERİLER

Yapılan çalışmada, Denizli kenti 2004, 2005 ve 2006 yıllarına ait trafik kaza verisinin kümelenerek kara noktaların ve bu noktalarda hangi zamanlarda daha çok hangi tür kazaların meydana geldiğinin belirlenmesi amaçlanmıştır. Bu amaçla, trafik kazalarının meydana geldiği noktaların koordinat verisi k-ortalamlar ve bulanık kümeleme analizine tabi tutularak, küme merkezlerine denk gelen kara noktalar belirlenmiştir. Tablo 5.20 kara noktalar olarak göze alınabilecek, kazaların en çok yoğunlaştığı noktaları göstermektedir.

İlk kara nokta olarak belirlenen, Gazi Mustafa Kemal Bulvarı ile İzmir Bulvarının kesiştiği İstasyon Kavşağı ve civarı, trafik yoğunluğunun fazla olması ve otobüs terminalinin yakınlığı açısından önem kazanmaktadır. Trafik yoğunluğunun pik saatlerde maksimuma ulaştığı bu kavşakta meydana gelen oluşumuna göre kaza türleri ise sırasıyla çoktan aza doğru;

- Yandan çarpma
- Arkadan çarpma
- Duran araca çarpma
- Sabit cisme çarpma
- Devrilme

şeklindedir.

Yandan çarpma türü kazalar daha çok sürücülerin kurallara uymaması, yani trafik ışığının yeşil yanmasını beklemeden hareket etmeleri veya kırmızı ışık yandığı halde yollarına devam etmek istemeleri sonucu meydana gelmektedir. Bu konuda toplumumuzun duyarlılığı trafik güvenliği lehinde yeterli değildir. Sürücü, yolcu, yaya, yetkili veya yetkisiz herkesin, yani halkın eğitilmesi, bilgilendirilmesi ve motive edilmesi çok büyük önem arz etmektedir. Bunun haricinde bu konudaki denetim ve yasal düzenleme faaliyetlerinin geliştirilmesinin (Kural ihlallerini önlemeye yönelik dürüst, sık ve teknolojidten yararlanılarak trafik kontrolleri ile trafik güvenliğini geliştirici ve uygulanabilir yasal düzenlemelerin yapılması v.b.) bu tür kazaların azalmasında çok büyük bir etkisi olacaktır.

Arkadan çarpma türü kazalar bakımından da İstasyon Kavşağı ilk sırada yer almaktadır. Bu tür kazalar genel olarak görünürlüğün yeterli olmaması, trafik yoğunluğu, öndeki aracı geçme isteği, öndeki taşıt sürücüsünün olumsuz davranışı, aceleci ve telaşlı davranışlar, sinyalizasyonlu kavşaklarda ışığın yeşilden kırmızıya dönmesi esnasındaki ani durmalar, yeterli takip mesafesinin sağlanmaması ve yüksek hız sonucu meydana gelmektedir. Buna ek olarak, araç lastiklerinin değiştirilmesi gereken zamandan daha uzun süre kullanımı, yol kaplamasının sürtünme direncinin (teker-kaplama sürtünme direncinin) yeterli seviyede olmaması ve araç fren sisteminin teknolojiye uygun olmaması ani durmalarda olumsuz etki yarattığından dolayı bu tür kazalara sebebiyet verebilmektedir. Bu tür kazaların önlenmesi için, araçların, yol ve çevresinin güvenli tasarlanması, teknolojiden mümkün olduğunca faydalanılması, yakın araç takibinin tespit edilerek gerekli cezai uygulamanın yerine getirilmesi alınabilecek tedbirler arasındadır.

Bu çalışmada sonucunda, Denizli kentinin diğer bir önemli kavşağı olan ve trafik yoğunluğunun çok fazla olduğu Çınar Meydanı, ikinci kara nokta olarak belirlenmiştir. Bu kavşak, şehrin günlük trafiğinin büyük çoğunluğuna hizmet etmektedir. Bu kavşakta meydana gelen oluşumuna göre kaza türleri ise sırasıyla çoktan aza doğru;

- Arkadan çarpma
- Duran araca çarpma
- Sabit cisme çarpma
- Yayaya çarpma

Bu kavşakta kaza sayısının fazla olması, sinyal zamanlamasındaki problemlerden, kavşak geometrik tasarımından, yaya trafiğinin çok fazla olmasından, bazı sürücülerin agresif araç sürmelerinden ve dikkatsizlikten dolayı kaynaklanabilir. Burada, kavşak sinyalizasyonun tekrar tasarlanması, teknolojiden mümkün olduğunca faydalanılarak kontrollerin yapılması ve gerekli cezai uygulamalarının yerine getirilmesi ile ileride olabilecek kaza sayısında ciddi bir azalma sağlanabilir.

Aslına bakılırsa bütün bu kara noktalarda meydana gelen kazaların ana nedenleri genel olarak aynıdır. Bu nedenlerden bazıları;

- Yüksek trafik yoğunluğu
- Trafik mühendisliği hizmetlerinde eksikliklerimizin olması (Optimize edilmemiş kavşak sinyalizasyonları, yanlış yol ve kavşak geometrik tasarımları)
- Toplumun trafik güvenliği konusunda yeterince duyarlı olmaması (Aşırı hız, agresif araç kullanımı, uygunsuz şerit kullanımları ve şerit değiştirmeleri v.b.)
- Yük ve yolcu taşımacılığında karayoluna ağırlık verilmesi,
- Trafik güvenliği konusunda gerek okul içi gerekse okul dışı eğitimimizin yetersiz olması,
- Trafik denetimi konusunda eksikliklerimizin olması

şeklinde sıralanabilir.

Bu aşamada trafik güvenliğinin sağlanabilmesi için alınabilecek tedbirler ise;

- Trafik mühendisliği faaliyetleri (Uluslararası tasarım standartlarını göz önüne alarak yol ve çevresinin güvenli tasarlanması, teknolojiden mümkün olduğunca faydalanılması, bölünmüş yol yapımı, otopark, oyun alanı, yaya kaldırımı, üst ve alt geçit inşaatı, güvenli taşıt üretilmesi, trafik kontrol merkezleri ile trafik düzenlenmesi, vb.)
- Eğitim faaliyetleri (Sürücü, yolcu, yayalar, trafikle yetkili yetkisiz herkesin, yani halkın eğitilmesi, bilgilendirilmesi, motive edilmesi, v.b.)
- Denetim ve yasal düzenleme faaliyetleri (Teknolojiden faydalanarak trafik kural ihlallerini önlemeye yönelik gerekli trafik kontrollerin sık olarak yapılması ve yasal düzenlemelerin caydırıcı olacak şekilde yeniden düzenlenmesi, vb.)

- Kombine taşımacılığının (kara, deniz, hava ve demir yolunun dengeli ve gerekli şekilde kullanılmasının) uygulamasının takip edilmesi.

şeklinde sıralanabilir.

Analizlerin sonucunda, kazaların büyük bölümünün yerleşim yerlerinde, bölünmemiş ve asfalt kaplamalı yollarda, genel olarak açık havalarda ve gündüzleri meydana geldiği belirlenmiştir. Bunun sebebi ise sürücülerin gündüzleri ve açık havalarda daha dikkatsiz olmalarıdır. Açık havalarda gündüzleri de farların yakılması, karşı taraftaki sürücü tarafından fark edilmeyi arttıracığından, kaza riskini azaltabilir. Kazaların büyük çoğunluğunun asfalt kaplamalı yollarda meydana gelmesinin sebebi ise, Denizli’ de kentiçi ve kent dışı yolların büyük çoğunluğunun asfalt kaplamalı yol olması olabilir. Yol kaplama yüzey düzgünlüğü sürüş konforundan ama sürtünmesi ise yol güvenliğinden sorumludur (Tunç 2004). Yeterli sürüş konforu olmayan bir yolda sürücü daha çabuk yorulmakta ve teker izi teşekküllü, deformasyon, çukur, ondülasyon, yüksek yamalar v.b. kaplama kusurları sürüş kontrolünü etkileyerek kazaların oluşmasına neden olmaktadır. Yol kaplama yüzeyinin sürüş güvenliği ise teker/kaplama sürtünme direncini temsil etmekte ve sürüş dinamik etkisi üzerinde en etkin rolü oynamaktadır. Kaplamadan beklenen sürüş güvenliği için;

- Kaplama direnci
- Teker izi olukları
- Yüzeysel drenaj
- Kaplama yüzeyinin ışığı az veya çok yansıtması

gibi faktörler dikkate alınmalıdır.

Oluşumuna göre kaza türleri göz önüne alındığında ortaya çıkan kara noktalar ise Tablo 6.1’ deki gibidir;

Tablo 6.1 Kaza türlerine göre tespit edilen kara noktalar

Kara Nokta veya Bölge	Oluşumuna Göre Kaza Türlerinde Görülme Sayısı
Sevindik	8
Yeni Adliye	8
İstasyon Kavşağı	7
Demokrasi Meydanı	5
Çınar Meydanı	4
Fatih Kavşağı	4
Ulusal Egemenlik Meydanı	4
Özay Gönlüm Meydanı	3
İzmir Bulvarı (Zafer Katrancı Köprülü Kav.)	3
Lise	3
Deliktaş Kavşağı	2
25. Sokak	2
Örnek Caddesi	2
29 Ekim Bulvarı	2
Devlet Hastanesi	2
Kıbrıs Şehitler Caddesi	2
Fatih Caddesi	2
Vilayet Önu Kavşak	1
Atatürk Caddesi	1
Dumlupınar Caddesi	1
Dokuzkavaklar Altgeçiti	1
Fevzi Çakmak Bulvarı	1
Meclis Caddesi	1
Yeşilköy Caddesi	1
201. Sokak	1

Buna göre Sevindik ve Yeni Adliye' de 11 oluşumuna göre kaza türünden 8 i görülmüş, İstasyon Kavşağı ve Demokrasi sırasıyla 7 ve 5' i görülmüştür. Bu şekilde çok sayıda kaza türünün meydana geldiği bu noktaları kesin kara nokta, buna nazaran daha çok tek tip veya 2 tip kaza türünün meydana geldiği diğer noktaları muhtemel kara noktalar olarak değerlendirmek doğru olacaktır.

Yukarıda detaylı biçimde açıklanan kümeleme analizleri sonucunda Denizli kent merkezindeki kara noktalar tespit edilmiştir. Fakat 2007 yılında Denizli kent merkezindeki bazı kavşakların geometrileri değiştirilmiştir. Bu değişimin gerçekleştirildiği ve kara nokta özelliği taşıyan kavşaklar ile ilgili olarak adrese dayalı 2007 yılı toplam kaza değerleri Denizli Emniyet Müdürlüğü Trafik Denetleme Şube

Müdürlüğü'nden elde edilmiştir. Söz konusu kaza değerleri 2004 - 2006 yılları arasında elde edilen adrese dayalı kaza değerleri ile Tablo 6.2'de karşılaştırılmıştır.

Tablo 6.2 Geometrisi değişen kavşaklardaki yıllık kaza sayıları

Kavşak Adı	Toplam Kaza Sayısı			
	2004	2005	2006	2007
İstasyon	53	162	226	89
Deliktaş	46	67	53	124
Bakırlı	31	14	26	121
Sevindik	17	28	41	87
YSE	26	23	31	48

Tablo 6.2' de görüldüğü üzere 2007 yılında geometrisi değiştirilerek kısmen farklı düzeyli biçimine (köprülü) dönüştürülen kavşaklardan İstasyon Kavşağı dışındakilerde kaza sayıları azalmamıştır. Dolayısıyla tespit edilen kesimler kara nokta özelliğini korumaktadır.

KAYNAKLAR

- Avcı, U. (2005), Bulanık Kümeleme, Yüksek Lisans Tezi, **Ege Üniversitesi Fen Bilimleri Enstitüsü**, İzmir, 157s
- Bezdek, J. C. (1973), Fuzzy Mathematics in Pattern Classification, PhD Thesis, Applied Math. Center, **Cornell University**, New York
- Bezdek, J.C., Pal, S.K. (1992), Fuzzy Models For Pattern Recognition: Methods That Search For Structures In Data, **IEEE Press**, New York, 539s
- Cebeci, Z. (2004), Türk Üniversitelerinin Bilimsel Yayın Sayıları İle Web’ te Bilgi Yayma Güçleri Arasındaki İlişkiler, **Akademik Bilişim 2007**, Kütahya
- Dushoff, J., Plotkin, J. B. and Levin, S. A. (2003), Hemagglutinin Sequence Clusters and The Antigenic Evolution of Influenza A Virus, **PNAS**, 99 (9): 6263-6268
- Espona, I. G., Gomez, J. T. and Carmona, J. B. (1995), Cluster Analysis Application To Class I Malocclusion, **The European Journal of Orthodontics**, 17 (3): 231-240
- Ester, M., Kriegel, H. P., Sander, J. and Xu, X. (1996), A Density-Based Algorithm For Discovering Clusters in Large Spatial Databases with Noise, **Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining**, Portland, 226-231
- Geurts, K. and Wets, G. (2003), Black Spot Analysis Methods: Literature Review, **Steunpunt Verkeersveiligheid Bij Stijgende Mobiliteit, RA-2003-07**, Diepenbeek, 30s
- Güler, N. (2006), Kümeleme Analizi, Yüksek Lisans Tezi, **Muğla Üniversitesi Fen Bilimleri Enstitüsü**, Muğla, 157s
- Hamarat, B. (1998), Türkiye’de Sağlık Açısından Homojen İl gruplarının Belirlenmesine İlişkin İstatiksel Bir Yaklaşım, Y. Lisans Tezi, **Anadolu Üniversitesi Fen Bilimleri Enstitüsü**, Eskişehir, 75s
- Han, J. and Kamber, M. (2000), Data Mining Concepts and Techniques, **Morgan Kaufmann Publishers**, San Francisco, 550s
- Huang, H. C., Nagaswamy, U. and Fox, G. E. (2004), The Application of Cluster Analysis In The Intercomparison of Loop Structures In RNA, **RNA**, 11 (4): 412-423
- Höppner, F., Klawonn, F., Rudolf, K. and Runkler, T. (1999), Fuzzy Cluster Analysis, **Wiley Publishers**, Wolfenbüttel, 300s
- Karpat, G. ve Yılmaz, V. (1997), Türkiye’deki Trafik Kazaları Oluş Şekillerinin, Kazanın Olduğu Yerdeki Trafik, Aydınlatma ve Yol Durumuna Göre İller Bazında İncelenmesi, **Uluslararası Trafik ve Yol Güvenliği Kongresi**, Ankara, 1-9

- Karaşahin, M. ve Saphioğlu, M. (2005), Coğrafi Bilgi Sistemi Yardımı İle Isparta İli Kentiçi Trafik Kaza Analizi, **Mühendislik Bilimleri Dergisi**, 12 (3): 321-332
- Kowtanapanich, W., Tanaboriboon, Y. and Chadbunchachai, W. (2005), Applying Public Participation Approach To Black Spot Identification Process, **IATSS Res (Int Assoc Traffic Saf Sci)**, 30 (1): 73-85
- Martin, S. (2001), Cluster Analysis for Web Site Organisation, **ITG Publication**, http://www.internettg.org/newsletter/dec99/cluster_analysis.html (12.12.1999)
- Murat, Y. S., Firat, M., Altun, S. (2008), Analysis of Traffic Accidents Using Fuzzy Clustering and Geographical Information Systems, **ID 639. AATT 2008 Conference Athens**, 28 - 30 May 2008
- Nasibov, E. N. (2005), Aggregation of Fuzzy Information on the Basis of Decompositional Representation, **Journal of Cybernetics and Systems Analysis**, 40: 309-318
- Sharma, S. (1996), Applied Multivariate Techniques, **John Wiley & Sons, Inc.**, New York, 493s
- Sheikholeslami, G., Chatterjee, S. and Zhang, A. (1998), WaveCluster: A Multire Solution Clustering Approach for Very Large Spatial Databases, **In Proceedings of the 24th VLDB Conference**, New York, 428-439
- Sintas, A. F., Cadenas, J. M. and Martin, F. (1999), Membership Functions In The Fuzzy C-Means Algorithm, **Elsevier North-Holland, Inc.**, 101: 49-58
- Şahin, M. ve Hamarat, B. (2002), G10 – Avrupa Birliği ve OECD Ülkelerinin Sosyo-Ekonomik Benzerliklerinin Fuzzy Kümeleme Analizi İle Belirlenmesi, **ODTÜ Uluslararası Ekonomi Kongresi VI**, Ankara
- Özdamar, K. (1997), Paket Programlar ile İstatistiksel Veri Analizi-2, **Kaan Kitapevi**, Eskişehir, 528s
- Tatlıdil, H. (2002), Uygulamalı Çok Değişkenli İstatistiksel Analiz, **Akademi Matbaası**, Ankara, 424s
- Tunç, A. (2004), Yol Güvenlik Mühendisliği ve Uygulamaları, **Asil Yayınları**, Ankara, 528s
- Turanlı, M., Özden, Ü. H. ve Türed, S. (2006), Avrupa Birliği' ne Aday ve Üye Ülkelerin Ekonomik Benzerliklerinin Kümeleme Analizi ile İncelenmesi, **İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi**, 9 : 95-108
- Ward, J. H. (1963), [Hierarchical Grouping to Optimize An Objective Function](#), **Journal of American Statistical Association**, 58: 236-244

WEB_1. (2008), Wikipedia, http://en.wikipedia.org/wiki/Cluster_analysis (06.02.2008)

Yıldırım, H. (1994), 'Coğrafi Bilgi Sistemleri', **MAM Uzay Teknolojileri Bölümü Yayını, MAM-Tn 04**, Marmara Araştırma Merkezi

ÖZGEÇMİŞ

02.09.1981 tarihinde doğan Alper ŞEKERLER, lise öğrenimini 1999' da Aydın Süper Lisesi' nde tamamladıktan sonra 2000 yılında Pamukkale Üniversitesi İnşaat Mühendisliği Bölümü' ne girmeye hak kazanmıştır. 2004 yılında İnşaat Mühendisliği Bölümü' nden mezun olduktan sonra 2005 yılının Eylül ayında Pamukkale Üniversitesi Fen Bilimleri Enstitüsü İnşaat Mühendisliği Ana Bilim Dalında Yüksek Lisans öğrenimine başlamıştır. Yüksek Lisans kapsamında 2006-2007 öğrenim yılında İtalya'nın Cenova kentinde bulunan Cenova Üniversitesi' nde ([Università degli Studi di Genova](#)) öğrenim görmüştür.