

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**HİYERARŞİK VE K-ORTALAMALAR YÖNTEMLERİYLE GRİD
NOKTALARININ KÜMELENMESİ**

Abdullah KIRMIZIBİBER

YÜKSEK LİSANS TEZİ

Harita Mühendisliği Anabilim Dalı

Uzaktan Algılama ve CBS Programı

Danışman

Prof. Dr. Türkay GÖKGÖZ

Kasım, 2019

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

HİYERARŞİK VE K-ORTALAMALAR YÖNTEMLERİYLE GRİD
NOKTALARININ KÜMELENMESİ

Abdullah KIRMIZIBİBER tarafından hazırlanan tez çalışması 25.12.2019 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Harita Mühendisliği Anabilim Dalı, Uzaktan Algılama ve CBS Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Türkay GÖKGÖZ

Yıldız Teknik Üniversitesi

Danışman

Jüri Üyeleri

Prof. Dr. Türkay GÖKGÖZ, Danışman

Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Alper ŞEN, Üye

Yıldız Teknik Üniversitesi

Doç. Dr. Ahmet Özgür DOĞRU, Üye

Yıldız Teknik Üniversitesi

Danışmanım Prof. Dr. Türkay GÖKGÖZ sorumluluğunda tarafımca hazırlanan Hiyerarşik ve K-Ortalamalar Yöntemleriyle Grid Noktalarının Kümelenmesi başlıklı çalışmada diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim

Abdullah KIRMIZIBİBER

İmza

TEŞEKKÜR

Yüksek lisans tez çalışmamda ve araştırmalarımın her safhasında bilgi, öneri ve tecrübelerini benden esirgmeden paylaşan, gerek lisansüstü çalışmalarında gerekse akademik faaliyetlerimde karşılaştığım her sıkıntıda yanımda olan, yüksek lisans eğitimim sırasında bilgileriyle ufkumu sürekli geliştiren, çalışma ahlakıma yön veren, hayatımın sonuna kadar hep bilgilerinden ve tecrübelerinden faydalanmak istediğim Yıldız Teknik Üniversitesi Harita Mühendisliği Bölümü Kartografya Anabilim Dalı Başkanı ve tez danışmanım Sayın Prof. Dr. Türkay GÖKGÖZ'e, bugünlere gelmemde en büyük emeği veren ve hayat mücadelesinde her sıkıntıda yanımda olan babam ve akıl hocam Seyfullah KIRMIZIBİBER'e, fedakarlığın vücut bulmuş hali annem Naime KIRMIZIBİBER'e, kardeşlerime, çalışmalarım sırasında bana her türlü yardım ve desteği sağlayan sayın Dr. Öğr. Üyesi Çağrı KARAMAN'a, Araş. Gör. Furkan Hamza HARPUTLUGİL'e ve bu süreçte her türlü kahrımı çeken kardeşimden öte yakınlık hissettiğim kıymetli insan Taha ÖZTÜRKÇÜ'ye en kalbî duygularla teşekkür ederim.

Abdullah KIRMIZIBİBER

İÇİNDEKİLER

KISALTMA LİSTESİ	vi
ŞEKİL LİSTESİ	vii
TABLO LİSTESİ	viii
ÖZET	ix
ABSTRACT	xi
1 GİRİŞ	1
1.1 Literatür Özeti	2
1.2 Tezin Amacı	3
1.3 Hipotez	3
2 YAPAY ZEKÂ	4
2.1 Yapay Zekâ Problemleri	5
2.2 Öğrenme	6
2.2.1 Bilgisayarlı Öğrenme	8
3 YÖNTEM	10
3.1 Kümeleme Analizi	10
3.1.1 Kümeleme Analizinin Amaçları	12
3.1.2 Kümeleme Analizinde Kullanılan Uzaklık ve Benzerlik Ölçütleri .	14
3.1.3 Kümeleme Analizi Yöntemleri.	17
3.1.4 Benzerlik Ölçümleri.	27
4 UYGULAMA	28
4.1 Veriler.	28
4.2 Benzerlik Ölçütleri: Öznitelikler.	29
4.2.1 En Düşük ve En Yüksek Kotlu Nokta	29
4.2.2 Ekstrem Noktalar	29
4.2.3 Eşit Ağırlıklı Öznitelikler.	31
4.2.4 Ağırlıklandırılmış Öznitelikler.	32
4.3 Hiyerarşik Yöntemle Kümeleme.	33

4.3.1	Eşit Ağırlıklı Öznitelikler İle Kümeleme	33
4.3.2	Ağırlıklandırılmış Öznitelikler İle Kümeleme	35
4.4	K-Ortalamalar Yöntemiyle Kümeleme	38
4.4.1	Eşit Ağırlıklı Öznitelikler İle Kümeleme	39
4.4.2	Ağırlıklandırılmış Öznitelikler İle Kümeleme	40
5	SONUÇLAR VE ÖNERİLER	41
	KAYNAKÇA	44
	TEZDEN ÜRETİLMİŞ YAYINLAR	48



KISALTMA LİSTESİ

B	Batı
D	Doğu
DEM	Digital Elevation Model
ED	En Düşük Kotlu
EY	En Yüksek Kotlu
G	Güney
GB	Güney Batı
GD	Güney Doğu
HKT	Hata Kareler Toplamı
K	Kuzey
KB	Kuzey Batı
KD	Kuzey Doğu
LIDAR	Laser Imaging Detection and Ranging
LOD	Ayrıntı Yüzeyi
RADAR	Radio Detection and Ranging
SPSS	Statistical Package for the Social Sciences
SYM	Sayısal Yükseklik Modeli
TIN	Triangulated Irregular Network
YSA	Yapay Sinir Ağları

ŞEKİL LİSTESİ

Şekil 2.1 Bilgisayarlı Öğrenme	9
Şekil 3.1 Kümeleme Yapısı.	11
Şekil 3.2 Ayırıcı ve Birleştirici Hiyerarşik Kümelemenin {a,b,c,d,e} veri nesneleri üzerinde gösterimi	19
Şekil 3.3 {a,b,c,d,e} Veri Nesnelerinin Hiyerarşik Kümelenmesi Sürecini Gösteren Dendrogram Gösterimi	20
Şekil 3.4 K-ortalama Yöntemi ile Kümeleme	26
Şekil 4.1 Uygulamada kullanılan SYM	28
Şekil 4.2 Örnek Grid Noktaları ve Yükseklik Değerleri.	30
Şekil 4.3 Profiller ve Profiller Boyunca Belirlenen Ekstrem Noktaları.	30
Şekil 4.4 Profiller Boyunca Belirlenen Ekstrem Noktaları.	31
Şekil 4.5 Eşit Ağırlıklı Öznitelikler ile Hiyerarşik Kümeleme Dendogramı	34
Şekil 4.6 Eşit Ağırlıklı Öznitelikler ile Hiyerarşik Kümeleme Sonucu Seçilen Grid Noktaları.	35
Şekil 4.7 Ağırlıklandırılmış Öznitelikler ile Hiyerarşik Kümeleme Dendogramı	37
Şekil 4.8 Ağırlıklandırılmış Öznitelikler ile Hiyerarşik Kümeleme Sonucu Seçilen Grid Noktaları.	38
Şekil 4.9 Eşit Ağırlıklı Öznitelikler ile K-Ortalamalar Yöntemi ile Yapılan Kümeleme Analizi Sonucu Seçilen Grid Noktaları	39
Şekil 4.10 Ağırlıklandırılmış Öznitelikler ile K-Ortalamalar Yöntemi ile Yapılan Kümeleme Analizi Sonucu Seçilen Grid Noktaları	40

TABLO LİSTESİ

Tablo 3.1 Kümeleme Yöntemleri	18
Tablo 4.1 Grid Noktalarının Öznitelikleri	29
Tablo 4.2 Grid Noktalarının Eşit Ağırlıklı Öznitelikleri	32
Tablo 4.3 Grid Noktalarının Ağırlıklandırılmış Öznitelikleri	33
Tablo 5.1 Uygulamaların Başlıca Sonuçları	42



Hiyerarşik ve K-Ortalamalar Yöntemleriyle Grid Noktalarının Kümelenmesi

Abdullah KIRMIZIBİBER

Harita Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Prof. Dr. Türkay GÖKGÖZ

Sayısal yükseklik modelleri (SYM) başlıca iki temel üzerine inşa edilmektedir: Grid ve TIN (Triangulated Irregular Network). Kimi durumlarda daha özet SYM'lere gereksinim duyulur. Daha özet SYM'ler genelleştirme yoluyla elde edilir. Grid temelli SYM'ler önemli noktaları seçilerek, diğerleri elenerek genelleştirilir. Bu çalışmada, grid temelli SYM'lerin genelleştirilmesi kapsamında, önemli grid noktalarının seçilmesi, bir yapay zekâ problemi olarak ele alınmış ve kümeleme analizi yöntemlerinin kullanılabilirliği araştırılmıştır.

Kümeleme, verilerin birbirine benzerliklerine göre gruplaştırılması işlemidir. Bu çalışmada, grid noktalarının birbirine benzerliklerini belirlemede kullanılmak üzere her bir grid noktasında şu soruların cevapları aranmıştır: En yüksek ya da en düşük kotlu nokta mıdır? Noktadan geçen dört profil (Doğu-Batı, Kuzey-Güney, Kuzeydoğu- Güneybatı, Kuzeybatı- Güneydoğu) boyunca kendisinden bir önceki ve bir sonraki noktadan daha düşük ya da daha yüksek kotlu bir nokta (ekstrem nokta) mıdır? Sekiz komşusu (Kuzey, Güney, Doğu, Batı, Kuzeybatı, Kuzeydoğu, Güneybatı, Güneydoğu) arasında ekstrem nokta(lar) var mıdır? Böylece, her bir grid noktasının on üç özniteliği belirlenmiştir. Grid noktaları bu öznitelikler

kullanılarak Hiyerarşik ve K-Ortalamalar yöntemlerine göre kümelenmiştir. Ekstrem noktaların kümelere dağılımı incelenerek, ekstrem noktaların belli kümelerde toplanıp toplanmadığı ve dolayısıyla grid noktalarının “Seçilenler” ve “Elenenler” olarak iki kümede toplanıp toplanamayacağı belirlenmeye çalışılmıştır.

Her iki yöneme göre, eşit ağırlıklı öznitelikler ve ağırlıklandırılmış öznitelikler kullanılarak, ikişer uygulama yapılmıştır. Uygulamalarda 1681 adet grid noktasından (41 satır, 41 sütun) meydana gelen bir SYM’nin 160 sınır noktası hariç 1521 noktası kullanılmıştır.

Sonuç olarak, ağırlıklandırılmış öznitelikler kullanılarak hiyerarşik kümeleme yöntemine göre yapılan uygulamada beş küme elde edilmiş olmasına karşın, ekstrem olmayan noktaların tamamının bir kümede toplanmış olması, bu kümenin “Elenenler Kümesi” olarak, diğerlerinin de “Seçilenler Kümesi” olarak değerlendirilebileceği görülmüştür.

Anahtar Kelimeler: Sayısal Yükseklik Modelleri, Kümeleme, Hiyerarşik, K-Ortalama

Clustering of Grid Points Using Hierarchical and K-Means Methods

Abdullah KIRMIZIBİBER

Department of Geomatics Engineering

Doctor of Philosophy Thesis

Advisor: Prof. Dr. Trkay GKGZ

Digital Elevation Models (DEM) are examined under two main headings: Grid and TIN (Triangulated Irregular Network). In some cases more summary DEMs are needed. More summary DEMs are obtained through generalization. Grid-based DEMs are generalized by selecting important points and eliminating others. In this study, the selection of important grid points is considered as an artificial intelligence problem and the usability of clustering analysis methods is investigated within the scope of generalization of grid-based DEMs.

Clustering is the process of grouping data according to their similarities. In this study, the answers to the following questions were searched in each grid point to determine the similarity of the grid points: Is it the highest or the lowest point? Is it a lower or higher elevation point (extreme point) along the four profiles passing through the point (West-East, South-North, Southwest-Northeast, Southeast-Northwest)? Are there extreme point (s) between eight neighbors (North, South, East, West, Northwest, Northeast, Southwest, Southeast)? Thus, thirteen attributes of each grid point were determined. Grid points are clustered according to Hierarchical and K-Means methods using these attributes. By examining the

distribution of the extreme points to the clusters, it was tried to determine whether the extreme points are collected in certain clusters and therefore the grid points can be collected in two clusters as “Selected and “Eliminated”.

According to both methods, equal weighted attributes and weighted attributes were used in two applications. In the applications, 1521 points were used except for 160 limit points of a DEM consisting of 1681 grid points (41 rows, 41 columns).

As a result, although five clusters were obtained by using hierarchical clustering method using weighted attributes, it was seen that all non-extreme points were collected in one cluster, and that this cluster could be considered as “Eliminated Cluster, and others could be considered as “Selected Cluster”.

Keywords: Digital Elevation Models, Clustering, Hierarchical, K-Means

Yeryüzü matematiksel olarak ifade edilmesi mümkün olmayan düzensiz üç boyutlu bir yüzeydir. Bu yüzeyi tam manasıyla birebir gösterebilmek için sonsuz sayıda noktaya ihtiyaç duyulmaktadır. Bu ise imkânsız olduğu için belirli sayıda nokta kümesi seçilir ve bu noktalar yardımıyla yeryüzü gösterilmeye çalışılır.

Sayısal Yükseklik Modeli (SYM) bir arazi yüzeyinin koordinatlarını ve yükseklik verilerini ifade eden üç boyutlu gösterimdir. Sadece koordinat bulmak ve yükseklik ölçümleri için kullanılabilen SYM aynı zamanda arazi eğimi, bakı, güneşlenme, rüzgar analizi vb. için de kullanılabilir (Makineci, 2016).

İlk olarak 1950'lerde ortaya atılan SYM'de veri toplama işi klasik yöntemlerle yapıliyordu. Veri elde etmek günümüze kıyasla o dönemlerde çok zor, uzun ve maliyetli bir işlemdi. Günümüzde kullanılan uydu görüntüleri, LIDAR (Laser Imaging Detection and Ranging), RADAR (Radio Detection and Ranging) teknolojileri sayesinde SYM üretmek çok daha hızlı, kolay ve maliyeti düşük bir hale gelmiştir. Bu sayede SYM üretimi oldukça yaygın kullanılan standart bir uygulama haline gelmiştir (Yastıklı ve Esirtgen, 2011).

Dünya yüzeyinin bir kısmının dijital temsili olarak, SYM'ler topografya veya diğer yüzeylerle ilgili olayları modellemek, analiz etmek ve görüntülemek için yaygın olarak kullanılmıştır. Fakat mevcut SYM verileri, kullanıcının farklı seviyelerdeki taleplerini her zaman karşılamaz ve teknolojinin farklı çözünürlüklerde yeni veriler üretmesini gerektirir. Bu işlem genellikle harita genelleştirmesi ile gerçekleştirilir. (Ai ve Li, 2010) Kullanıcı, bir yandan veri hacmini düşürmek, diğer yandan ana bilgileri (özellikle de veride gizlenmiş yapısal bilgiyi) korumak ister. SYM sadeleştirilmesi üzerinden örnek verecek olursak, en düşük seviyedeki işlem, arazi gösterimini daha basit ve yumuşak hale getirmek için raster hücrelerin yükselmesini veya kesilmesini sağlamaktır. Ancak, ana vadileri ve sırtları korumak ve yalnızca küçük olanları kaldırmak için coğrafi anlam düzeyinde karar verilir (Poorten ve Jones, 1999).

Haritaları, hava fotoğrafları ve uydu görüntülerinden ayıran en büyük özelliği haritası olduğu bölgenin bütün somut objelerini içermemesidir. İçereceği objeler kullanım amacına, konu ve ölçeği göz önüne alınarak önemli detayların seçilip geri kalanların elenmesi işlemine genelleştirme denir. Bu şekilde daha detaylı bir haritadan daha az ayrıntı içeren bir harita elde edilir. SYM'lerde genellikle birçoğu gereksiz olan çok fazla noktadan meydana gelmektedir. Genelleştirme işlemi yapılırken bu kümelerin bazılarının elenip bazılarının ise seçilmesi gerekir. En önemli grid noktalarının seçilmesi gerekmektedir. Burada, bir noktanın önem derecesini belirleyen kriter, o noktanın yüzey gösterimine katkısıdır. Bir noktanın yüzey gösterimine katkısı ne kadar fazla ise önem derecesi de o kadar fazladır (Chen ve Guevara, 1987).

1.1 Literatür Özeti

SYM genelleştirilmesi konusunda onlarca yıldan beri araştırmalar yapılmıştır. Yaygın olarak, yeniden örnekleme, düşük geçişli filtreleme ve dalgacık yöntemleri gibi görüntü işleme teknolojilerini kullanmaktır. Bir görüntüyü piksel sayısını azaltarak yeniden boyutlandırma tekniğine dayanan yeniden örnekleme, veri hacmini azaltmada etkilidir, ancak orijinal arazi yapısı özelliklerini korumada etkisizdir. Düşük geçişli filtreleme, SYM genelleştirilmesi veya TIN (Düzensiz Üçgen Ağı) basitleştirmesinde yaygın olarak kullanılır (Chen, 1989; Weibel, 1992). Bu yöntem, yüksek frekansları engelleyebilir ve ayrıntıları etkin bir şekilde kaldırmak ve SYM'yi daha yumuşak hale getirmek için düşük frekansların geçmesine izin verebilir. Ancak karakteristik ve karakteristik olmayan coğrafi özellikleri ayırt edemez. Genelleştirilmiş sonuç, bazı önemli özellikleri kaçınılmaz bir şekilde atar, örneğin vadi başı kesilebilir. Başka bir problem ise, filtreleme genelleştirilmesinin kritik olarak filtre penceresinin boyutuna bağlı olmasıdır (Ai ve Li, 2010). Weibel (1992) filtrelemeyi gerçekleştirmek için 3×3 veya 5×5 bir pencere uygulamaktadır. SYM filtreleme yöntemlerinin karşılaştırılmasından sonra, Schmid-Mcgibbon (1995), ortaya çıkan haritaların görsel yorumuyla birlikte, pencere boyutunun doku analizi temelinde seçilmesi gerektiğine vurgu yapmıştır. Analiz, varyansların arazi pürüzlülüğüne ve alan boyutuna göre farklılık gösterdiğini ve yüksek dereceli türevlerin arazi parametreleri için varyansın

arttığını göstermektedir (Schmid-McGibbon, 1995). Dalgacıklar, bazı gösterim ölçekleriyle eşleşmesi için verileri belirli çözünürlüklerde farklı frekans bileşenlerine bölen matematiksel işlevlerdir. Dalgacıklar, görüntü sıkıştırma işleminde yaygın olarak kullanılmaktadır ve ayrıca SYM genelleştirmesine de dahil edilmiştir (Wu, 2003; Bjørke ve Nilsen, 2003; Wälder ve Buchroithner, 2006). Bu doğrusal olmayan filtreleme tekniği, pürüzlü alanlardaki büyük dalgacık katsayılarının tutulması anlamında uyarlanırken, pürüzsüz alanları temsil eden eşik değerden küçük tüm katsayılar atılır (Ai ve Li, 2010).

Weibel (1992), arazi genelleştirilmesi için metodolojileri sistematik olarak incelemekte ve bunları üç tipe ayırmaktadır: Küresel filtre, seçici filtre ve sezgisel yaklaşım. Bu sınıflandırmaya göre, yeniden örnekleme ve düşük geçişli filtreleme, görüntü işlemeyen pozisyon değişmeyen filtreleme teknikleri ile ilgili olan küresel filtre tipine aittir.

1.2 Tezin Amacı

Grid temelli SYM'lerin genelleştirilmesi konusunda bugüne kadar birçok çalışma yapılmış olmasına rağmen konu hiçbirinde bir yapay zekâ problemi olarak ele alınmamıştır. Oysa ki grid temelli SYM'lerin genelleştirilmesinde özde yapılan önem derecesi yüksek olan grid noktalarının bir küme altında toplanmasıdır ve kümeleme amacı doğrultusunda geliştirilmiş çeşitli yapay zeka yöntemleri vardır.

Bu tezin amacı, grid noktalarının yaygın olarak kullanılan iki kümeleme yöntemiyle (hiyerarşik ve k-ortalamlar) “Seçilenler” ve “Elenenler” olmak üzere iki kümede toplanıp toplanamayacağının belirlenmesidir.

1.3 Hipotez

Grid temelli SYM'lerin genelleştirilmesi için elenecek ve seçilecek grid noktalarının belirlenmesi işleminde kümeleme yöntemlerinin kullanılması sonucunda, seçilecek grid noktalarının önem derecesi yüksek olan grid noktaları olabileceği öngörülmüştür. Ayrıca seçilecek grid noktalarını belirleme işleminin, kümeleme yöntemleriyle yapılması sayesinde çok daha hızlı, kolay ve doğru olacağı düşünülmektedir.

2 YAPAY ZEKÂ

Yapay zekâ, kabaca; genellikle insana özgü nitelikler olarak kabul gören anlam çıkartma, akıl yürütme, genelleme ve geçmiş deneyimler sayesinde öğrenme gibi yüksek zihinsel süreçlere ilişkin görevleri bir bilgisayarın veyahut bilgisayar denetimli bir makinenin, yerine getirebilme becerisi olarak tanımlanmaktadır (Nabiyev, 2016).

Yapay zekâ kavramının birçok farklı tanımı da bulunmaktadır.

Örneğin;

- Geneseresh ve Nilsson'a göre yapay zekâ, asıl hedefi, doğadaki varlıkların sergilemiş oldukları akıllı davranışlarını yapay olarak üretmeyi amaçlayarak, akıllı davranış üzerine yapılan bir çalışmadır (Charniak ve McDermot, 1985).
- "Yapay zekâ, bilgisayarlara insanların yaptıklarını yaptırabilme çalışmasıdır" (Pospelov, 1990).
- Ayr göre ise yapay zekâ, akıllı programları hedef alan bir bilimdir. Bu programlar aşağıdakileri gerçekleştirebilmelidir (Copeland, 1993).
 1. Karmaşık problemleri insan düşünmesini taklit ederek çözebilmek
 2. Bir durum karşısında kişiye cevap verebilmek yani yorumlarını açıklayabilmek ve
 3. Hem yeni bilgiler öğrenerek uzmanlığını geliştirmek hemde eski bilgilerini yenilerle uyumlu biçimde kullanarak bilgi tabanını genişletmek.

Yapay zekâ kavramının tarihsel gelişim süreci Pirim (2006) tarafından aşağıdaki gibi dönemlere ayrılmıştır.

Karanlık Dönem (1965-1970): Çok fazla gelişimin görülmediği bir dönemdir. Düşünen bir mekanizma geliştirmek isteyen bilgisayar uzmanları, bu sayede

sadece verileri yükleyerek akıllı bilgisayarlar yapmayı amaçladılar. Bekleme dönemi olarak da adlandırılabilir.

Rönesans Dönemi (1970-1975): Çok hızlı bir şekilde artacak olan gelişmelerin ilk adımlarının atılarak bugünkü yapay zekâ gelişmelerinin temelinin oluşturulduğu dönem olmuştur. Hastalık teşhisi gibi sistemler yapay zekâcılar tarafından geliştirdiler.

Ortaklık Dönemi (1975-1980): Diğer bilim dallarından da (dil ve psikoloji gibi) yapay zekâ araştırmacıları yararlanmaya başladılar.

Girişimcilik Dönemi (1980-?): Gerçek dünyanın ihtiyaçları göz önüne alınarak, yapay zekâ laboratuvarların dışına çıkarılarak, çok daha kompleks uygulamalarda düşünülmüştür. Günümüzde de devam etmekte olan dönemdir.

2.1 Yapay Zekâ Problemleri

Genel olarak yapay zekâ problemleri aşağıdaki gibi büyük sınıflara ayrılmaktadır (Nabiyev, 2016).

- Problem çözümü: Burada sezgisel yöntemlerle karmaşık kombinasyonel özellikli problemlerin çözümü ele alınmaktadır. Problem çözümünde insanın psikolojik açıdan gerçekleştirdiği süreçlerin modellenilmesi amaçlanır.
- Oyunların modellenmesi: Burada genellikle bilgisayarın insana benzer bir şekilde kararlar alabilmesinin nasıl sağlanacağına değinmek için go, dama, satranç gibi strateji oyunları ele alınmaktadır.
- Bilgilerin modellenmesi: Burada bilgilerin modellenmesi ve onların farklı yöntemlerle bilgisayara aktarılması üzerinde durulmaktadır. Ayrıca çok büyük boyuttaki bilgilere bilgisayarlarda farklı yöntemler kullanılarak nasıl hızlı erişim sağlanacağı açıklanmaktadır. Öğrenmeyi de içeren bulanık sistemler, yapay sinir ağları, anlamsal ağlar, çerçeveler farklı bilgi modellerini oluşturmaktadır. Bilgilerin modellenmesi, örüntü tanıma, doğal dil işleme, uzman sistemler ve diğer yapay zekâ problemlerinin çözümünde önem taşımaktadır.

- Otomatik teorem ispatı: Mantık ve matematik ile ilişkili bir şekilde önermelerin ispat edilmesi ve yenilerinin bulunması üzerinde durulmaktadır.
- Uzman sistemler: Uzmanlık gerektiren konulardaki problemlerin çözümü ile ilgili özel bilgileri içermektedir. Bu sistemler, kendi alanlarında uzmanlaşmış kişiler tarafından çözümlenen problemlerin bilgisayarda modellenmesi şeklinde tanımlanabilir. Uzman sistemlere örnek olarak tıbbi, askeri ve mühendislik alanlarındaki uygulamalar verilebilir.
- Doğal dilin işlenmesi: Genellikle soru-cevap biçiminde olan diyaloglu sistemler, anlamının modellenmesi, cümlelerin analizi (morfolojik, sentaks, semantik ve pragmatik), imla hatalarının algılanması ve düzeltilmesi, otomatik çeviri sistemleri ele alınmaktadır.
- Örüntü tanıma: Görsel ve işitsel nesnelerin tanınması araştırılmaktadır. Tıbbi bir görüntünün, biometrik özelliklerin tanınması, sahne (scene) analizi, el yazısı veya basılı karakter tanınması, bilgisayarlı dudak okuma veya yüz tanıma bu cinsten olan problemlerdendir.
- Robotik: Akıllıca denilebilen bilgilerle donatılmış ve mekanik biçimde tasarlanmış ve sistemler araştırılmaktadır.

2.2 Öğrenme

Öğrenme, bir sistemin çevresine daha iyi uyum sağlayabilmesi için kendi kendini yeniden düzenlemesi olarak adlandırılabilir. Öğrenme ile bireyler, ortamdaki değişimlere bağlı olarak, davranış yetilerini en iyi şekilde kullanarak yaşamlarını sürdürebilme imkanı sağlamaktadır (Nabiyev, 2016).

Psikologlara göre öğrenmenin genel olarak üç türü bulunmaktadır (Nabiyev, 2016):

- a. Alışkanlık
- b. Çağrışımlı öğrenme
 - Klasik koşullanma (şartlı refleks)
 - Edimsel koşullanma (yaparak şartlanma)
- c. Karmaşık öğrenme

Alışkanlık aynı davranış sürekli olarak yapıldığında düşünmeden yapılacak hale gelen alışlagelmiş ve ciddi sonuçlar doğurmayan en basit öğrenme şeklidir.

Çağrışimli öğrenemenin temeli, kişinin daha önceden tepki göstermediği herhangi bir davranışına sürekli ödül veya ceza verilmesi durumunda tepki göstermeye başlamasıdır.

Klasik koşullanma yoluyla öğrenme, Rus bir fizyolog olan I. Pavlov tarafından ilk olarak ortaya atılmıştır. Pavlov, köpeklerin sindirim sistemleriyle alakalı araştırma yaparken, bir köpeği ses geçirmeyen bir odaya koydu ve yiyecek vermeden hemen önce zili çalmaya başladı. Buradaki amaç yemek verme işlemiyle zil sesini eşleştirmektir. Bu işlem tekrar tekrar yapıldıktan sonra, zil çaldığı anda yemek verilmesi bile köpeğin salyalarının aktığını gözlemlenmiştir. Bir köpeğin yiyecek ağzına girdiği zaman fizyolojik olarak sindirimi başlatan salyayı salgılaması gerekirken, yiyecek ağzına girmeden yiyeceği hatta ve hatta yiyecek getiren kişiyi dahi gördüğünde salya salgıladığını gözlemlemiştir (Erden, Akman, 2001:133).

Pavlov'un çalışmasından da anlaşılacağı üzere birbiriyle bağlantılı olan uyarıcılar belirli tepkilerin ortaya çıkması yönünde tepkisel olarak koşullandırılabilir. Klasik koşullanmanın uyarıcı-tepki ilişkisine örnek olarak, Pavlov'un deneyinde kullandığı zil sesi ile salya arasında bulunan ilişki gösterilmektedir. Tepkisel koşullanmanın oluşabilmesi başlangıçta Pavlov'un deneyinde kullandığı yiyecek ve salya arasındaki gibi doğal uyarıcı tepki ilişkisinin bulunması gerekir. Daha sonra deneyde kullanılan zil sesi gibi koşullu uyarıcının yemek gibi koşulsuz uyarıcıdan hemen önce verilmesi ve bu iki uyarıcının salya salgılaması gibi beklenen tepki yönünden birleştirilmesi gerekir (Elden, 2003).

Edimsel koşullanma ile öğrenmeyi ise ceza ve ödül ikilisi oluşturur. Kişinin istenmeyen bir davranışı sergilemesi durumunda ceza verileceği, istenilen davranışı sergilemesi durumunda da ödül verileceği belirtilir. Polis eğitim köpeklerinin ve sirkteki hayvanların eğitimi sırasında kullanılan öğretme yöntemi budur (URL1). Amerikalı bir psikolog olan B.F. Skinner (1904-1990) edimsel koşullanma ile alakalı ilk deneyleri yaparken sadece tepkisel davranışların değil edimsel davranışların da şartlanabileceğine dikkati çekmiştir (URL2).

2.2.1 Bilgisayarlı Öğrenme

Öğrenme yöntemlerinden bağımsız olarak insan, yeni ve farklı bir olayla karşılaştığı zaman bilgi deposunu zenginleştirir. Örneğin çocuktan, yanan bir sobaya elini sürmemesini istiyorsunuz hatta görsel olarak mimiklerinizi de kullanarak bunu ifade ediyorsunuz. Bu onda daha merak oluşturuyor ve yanan sobaya yalnızca elini dokundurduktan sonra söylenenlerin ne demek olduğunu anlıyor. Çocuğa sobaya dokunmayı tekrar teklif ettiğiniz zaman, artık kesinlikle dokunmayacaktır. Bilgilerin bu şekilde elde edilmesine 'geri beslenmeli öğrenme' adı verilir. Bir bilgisayar programı da buna benzer şekilde çalışmaktadır. Var olan gerçekler temelinde mantıksal sonuçlandırma mekanizması da kullanarak yeni bilgilerin eklenmesiyle sistemin öğretilmesi gerçekleştirilir (Nabiyev, 2016).

Bir öğrenme sisteminin oluşturulması aşağıdaki şekildeki gibi özetlenebilir:

Sistem, sadece zıt bilgiyle karşılaştığında geri beslenme mekanizması yardımıyla tabandaki değerlerini değiştirir.

Sistem, girilen bilgi parametrelerini, daha önceden kaydedilmiş değerlerle karşılaştırır ve en yakın olan benzerliği arayarak onları eşleştirmeye çalışır. Tüm bu işlemleri, sistemin mantıksal çıkarım mekanizması üstlenmektedir (Nabiyev, 2016).

Öğrenmenin farklı türleri ve hiyerarşik bir yapısı vardır. Bilgisayarlı öğrenme de insanlarınkine benzer hiyerarşik yapıya sahiptir. (Şekil 2.1). Hiyerarşik yapı aşağıdaki gibi beş seviyeli bir yapıdır (Nabiyev, 2016).

Birinci seviyede en basit öğrenme türü olan programlanmış öğrenme söz konusudur. İlkel bilgisayar programları bu türdendir. Ardışık emirler neyin yapılması gerektiğini sıralı biçimde ifade etmektedir.

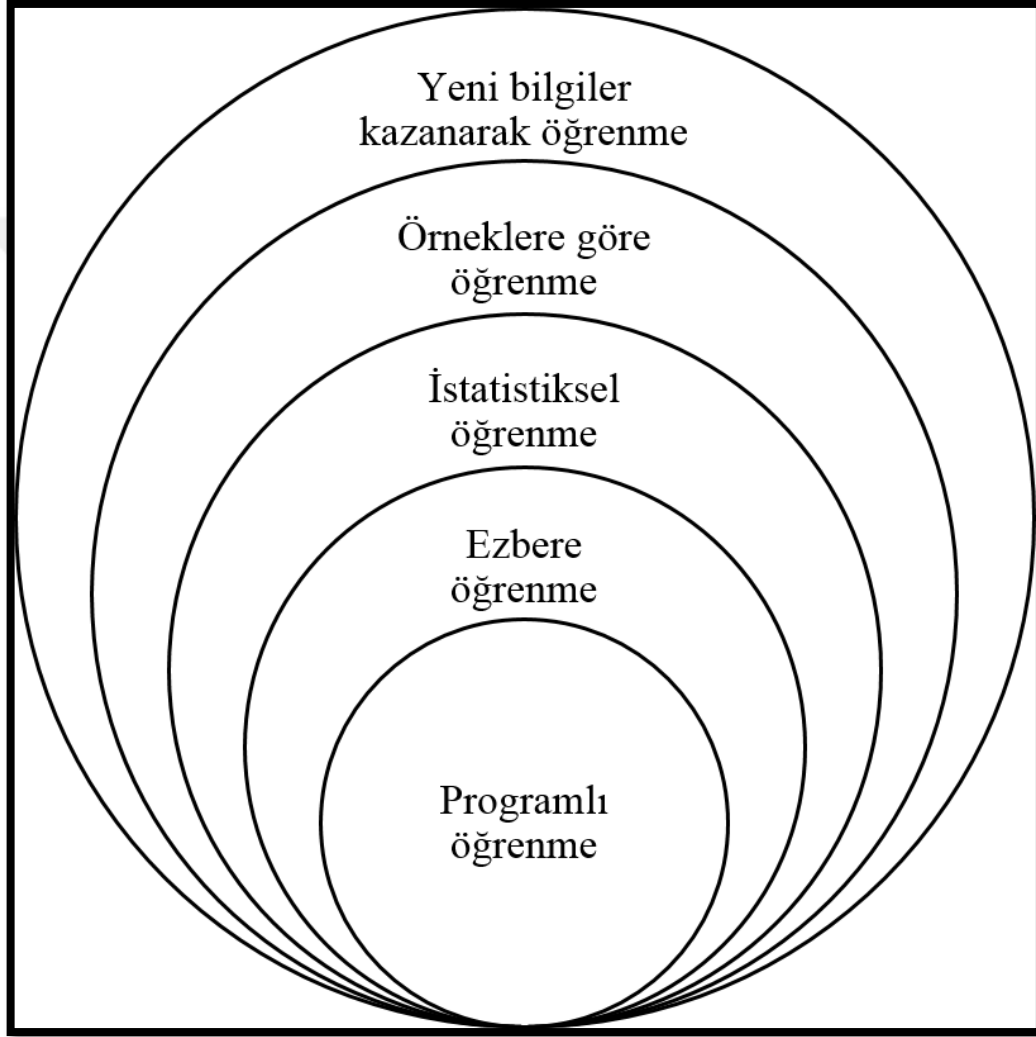
İkinci seviye, ezbere öğrenmeye karşılık gelir; her duruma ilişkin tepkiler bir bütün olarak tutulmaktadır.

Üçüncü seviyede, istatistiksel öğrenme söz konusudur. Burada sistemin birçok kez çalıştırılması (öğretilmesi) sonucu, tepkilere uygun olan bileşenler belirlenmektedir.

Dördüncü seviyede sistem, örneklere göre genelleştirme yapabilmekte ve öğretmenli öğrenme söz konusu olmaktadır.

Beşinci seviyede ise sistem, yeni hipotezler ve durumlar oluşturarak kendi kendine (öğretmensiz veya danışmansız) öğrenebilmektedir.

Bu tez çalışması, beşinci seviye öğrenme kapsamına giren kümeleme analizinin SYM genelleştirmesinde kullanılabilirliğinin araştırıldığı bir çalışmadır.



Şekil 2.1 Bilgisayarlı öğrenme türleri (Nabiyev, 2016)

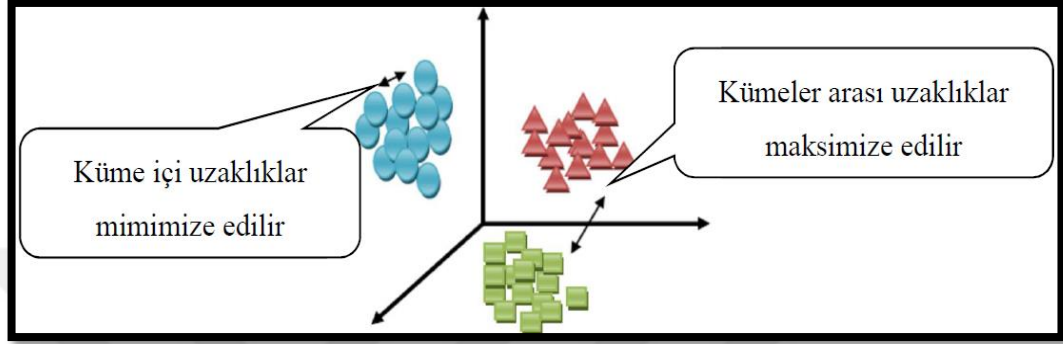
Bu tez çalışmasında, SYM genelleştirmesi kapsamında önemli grid noktalarının seçilmesi, bir yapay zekâ problemi olarak ele alınmış ve kümeleme analizi yöntemlerinin kullanılabilirliği araştırılmıştır. Grid noktalarının yükseklikleri ve komşuluk özellikleri esas alınarak toplam on üç adet öznitelik belirlenmiştir. Her bir yöntemle, eşit ağırlıklı öznitelikler ve ağırlıklandırılmış öznitelikler kullanılarak, ikişer uygulama yapılmıştır. Kümeleme analizi SPSS (Statistical Package for the Social Sciences) yazılımı yardımıyla, hiyerarşik kümeleme ve k ortalamar yöntemleriyle gerçekleştirilmiştir. Ekstrem grid noktaları AutoLISP programlama dilinde yazılmış olan programlar yardımıyla belirlenmiştir. Grid noktalarının komşuluk özellikleri ise AutoCAD ortamında manuel belirlenmiştir.

3.1 Kümeleme Analizi

Kümeleme analizi veri kümesindeki bilgileri veya değişkenleri belirlenen yakınlık kriterlerine göre gruplara ayırma işlemi olarak adlandırılmaktadır. Oluşturulan kümelerde, küme içindeki veriler arasındaki benzerlik olabildiğince fazla, kümeler arasındaki benzerlik ise olabildiğince az olmalıdır. Kümeleme analizinde verilerin dağıtılacağı sınıflar önceden belli olmayıp verilerin kaç farklı kümede toplanacağı, hangi verinin hangi kümeye atanacağı mevcut verilerin birbirlerine olan benzerliklerine göre belirlenir.

1939 yılında ilk kez kullanılan kümeleme analizi, 1960'lı yıllardan sonra yaygınlaşmıştır (Anderberg, 1973). Kümeleme analizi önceden belirlenen benzerlik kriterlerine göre birbirine çok benzeyen verileri aynı küme içerisinde toplar. Kümeleme işleminin sonucunda oluşan kümelerde kendi içindeki türdeşlik (homojenite) ve kümeler arasındaki heterojenlik olabildiğince en yüksek seviyededir. Yani, bir kümeyi oluşturan veriler birbirleriyle benzeşirken, diğer kümelerin verileriyle benzeşmeyecektir. Sonuçta sınıflandırma işlemi başarılı bir şekilde gerçekleştirildiyse tüm nesneler geometrik olarak işaretlendiğinde, aynı

küme içindeki nesneler birbirlerine oldukça yakın, farklı kümelerde bulunan nesneler ise birbirlerinden oldukça uzak olacaktırlar (Kalaycı, 2006). Eğer yapılan kümeleme işlemi başarılı olduysa, grafiksel gösterimde, aynı kümede yer alan nesneler birbirine oldukça yakın, farklı kümelerde yer alan nesneler ise birbirlerinden oldukça uzak olacaktır (Hair vd., 2006). Şekil 3.1’de başarılı bir kümeleme işlemi sonucunda oluşacak kümeleme yapısı gösterilmiştir.



Şekil 3.1 Kümeleme yapısı (Çağlar, 2017)

Kümeleme analizi, farklı yapılar halinde karşılaştığımız verilerin küme yapılarını ve sayılarını araştıran bir yöntemdir. Çok değişkenli istatistiksel yöntemlerden biri olan kümeleme analizi grupları net bir şekilde bilinmeyen; değişkenleri ve birimleri birbirlerine benzerlik oranları dikkate alınarak alt kümelere veya gruplara ayırır. Kümeleme analizinin birincil hedefi verileri sahip oldukları karakteristik özelliklerine göre gruplara ayırmaktır (Tatlıdil, 2002).

İstatistik çalışmalarında çok fazla veriye ihtiyaç vardır. Ancak verilerin bu kadar çok olması birçok istatistiksel yöntemin kullanılmasında problemler oluşturmaktadır. Bu yüzden yöntemlerin problemsiz bir şekilde kullanılabilirliğini sağlamak için verilerin daha az verilerden oluşan gruplara ayrılması gerekmektedir. Kümeleme analizinin amacı da verileri benzerliklerine göre gruplandırmaktır (Erilli 2009).

Kümeleme analizi birbirlerine en çok benzeyen verileri aynı grup altında toplaması amacıyla diskriminant analizine, birbirine en çok benzeyen değişkenlerin aynı gruplarda toplanması amacıyla faktör analizine benzer olup veri indirgeme özelliğine de sahiptir (Çakmak 1999: 187-205). Diskriminant analizi daha çok yeni bir gözlemin önceden bilinip hali hazırda bulunan bir gruba

uygun bir şekilde yerleştirilmesi amacıyla kullanılmakta ve geleceğe göre öngörülerde bulunulmasına izin vermektedir (Selanik 2007).

Kümeleme analizinin avantajları aşağıdaki gibi sıralanabilir (Kaufman ve Rousseeuw, 1990):

- İlişkilerin görüntülenmesi: Sonuçların grafik gösterim olarak verilmesi kümeleme analizinin en önemli özelliklerinden biridir. Çünkü görsel olarak verilen sonuçlardan faydalanılarak benzerlikler daha kolay tespit edilebilir.
 - Anormalliklerin tespiti: Grafiklerden benzerliklerin tespit edildiği gibi, aykırı değerler de kolaylıkla tespit edilebilir, bu sayede aykırı değerler ile ilgili aksiyonlar alınabilir.
 - Diğer veri madenciliği teknikleri için ön hazırlık: Karar ağaçları gibi bazı teknikler daha küçük veri yığınları üzerinde çalışabilir. Büyük veri yığınlarının bu tekniklerde doğru çalışması beklenemez. Kümeleme analizi kullanılarak büyük veri yığınları daha küçük veri yığınları halinde gruplara ayrıldığı için bu tür tekniklerin kullanılabilmesi için zemin hazırlanır.
- Bunların yanı sıra kümeleme analizinin de yetersiz olduğu bazı yönleri bulunmaktadır. Bunlar (Kaufman ve Rousseeuw, 1990);
- Sonuçların anlaşılması zordur: Uyulması gereken belli başlı kuralların bulunmaması tahminlerin belli bir hata payı ile yapılmasına sebep olmaktadır.
 - Farklı özellikler içeren nesnelerin karşılaştırılması zordur.

3.1.1 Kümeleme Analizinin Amaçları

Küçük boyutlu veri tabanlarındaki değişkenleri gözlemlemek, sınıflandırmak, aralarındaki ilişkiyi araştırmak, düzenlemek gibi bazı birtakım işlemlerin yapılması çok da zor bir süreç değildir. Fakat bu durum büyük boyutlu veri tabanları için aynı kolaylıkta değildir. Çünkü büyük veri tabanlarındaki değişkenlerin karmaşık halde dizilmiş olması, benzer özellikteki değişkenleri ayırt edememeye sebep olmaktadır. Bu nedenle, araştırmacılar bu gibi durumlar karşısında istatistiksel analiz yöntemlerinden biri olan kümeleme analizini kullanmaktadır (Alpar, 2013).

Kümeleme analizinin birincil hedefi, araştırma sonucunda elde edilen gözlemlerin birbirlerine olan benzerliklerine bakarak en az iki olmak şartıyla gruplara ayırmaktır. Kümeleme analizinin en temel kullanım sebebi araştırma amaçlıdır. Bunun yanı sıra objektif bir sınıflandırma geliştirmede de sıklıkla kullanılmaktadır. Kümeleme analizi işlemi neticesinde oluşturulan gruplar, nesnelerin yapılarına bağlı olarak varsayımlar oluşturmasını sağlayabilirler. Bir araştırma tekniği olarak kabul edilen kümeleme analizi, aynı zamanda kanıtlayıcı amaçlar için de kullanılır (Kalaycı, 2006).

Kümeleme analizinin Özdamar (1999)'a göre kullanım amaçları:

- n sayıda veriyi, p sayıda değişkene ilişkin özelliklerine göre, küme içerisinde olabildiğince benzer ve kümeler arası olabildiğince farklı olacak şekilde alt gruplara ayırmak,
- p sayıda değişkeni, n sayıda veride saptanan değerlere göre benzer özellikleri açıkladığı varsayılan alt kümelere ayırmak ve ortak etmen yapıları ortaya koymak,
- Verileri ve değişkenleri birlikte ele alarak n sayıda veriyi p sayıda değişkene göre benzer özelliklere sahip alt kümelere ayırmak,
- p değişkene göre saptanan değerler bakımından bireylerin biyolojik ve taksonomik sınıflamasını ortaya koymak.

Kümeleme analizinin Alpar (2011)'e göre ise kullanım amaçları:

- Kümeleme analizi bir veri içindeki doğal grupların tanımlanması amacıyla kullanılır.
- Çoğunlukla olmasına rağmen sadece gözlemleri kümelemek amacı ile kullanılmakla kalmayıp aynı zamanda değişkenleri kümelemek amacı ile de kullanılır. Ancak değişkenlerin kümelemesi için çoğunlukla faktör analizi kullanılır.
- Bazen gözlemler arasında nesnel olmayan yaklaşımlarla kümeler oluşturulabilmekte ya da nesnel olduğu düşünülen kümelendirmeler üzerinde tartışmalar yaşanabilmektedir. Bu gibi durumlarda, kümeleme analizi, soruna nesnellik getirebilmekte ve nesnel sınıflamalar yapılmasına yardımcı olabilmektedir.

- Boyut indirgemek amacı ile de kümeleme analizi kullanılabilir.
- p sayıda değişken kümeleme analizi yardımıyla incelendiği zaman aykırı yada aşırı değer olarak adlandırılacak gözlemler belirlenebilir.
- Kümeleme analizi, verinin yapısı hakkında öne sürülen ya da önceden belirlenen hipotezleri incelemek amacı ile de kullanılabilir.

Tatlıdil (1996) ise, bazı özel amaçlar için de kümeleme analizinin kullanılabileceğini düşünmüştür. Bunlar;

1. Gerçek tiplerin belirlenmesi,
2. Gruplar için ön tahmin,
3. Model uydurmanın kolaylaştırılması,
4. Uç değerlerin belirlenmesi,
5. Veri yapısının netleştirilmesi,
6. Hipotezlerin testi,
7. Veri indirgemesi (veriler yerine kümelerin değerlendirilmesi),

3.1.2 Kümeleme Analizinde Kullanılan Uzaklık ve Benzerlik Ölçütleri

Kümeleme analizinin asıl amacı gözlemlenen birey veya nesneler arasındaki benzerlik ve uzaklıkları belirlemektir. Kümeleme analizinde kullanılan bütün yöntemlerde amaç, küme içi benzerliği maksimum düzeyde, kümeler arası benzerliği ise minimum seviyede tutmaktır. Bu sebepten dolayı kümeleme analizinin doğruluğu seçilen uzaklık ölçütüne bağlıdır.

Uzaklık, veriler arasındaki zıtlık yada uyumsuzluğun boyutunu belirleyecek farklılıkları ölçer. Benzerlik ise iki veri veya iki özellik arasındaki ilişkinin kuvveti olarak açıklanır. Alınan ölçeğe veya veri tipine bağlı olarak bu nicel değer değişik yollardan elde edilir. Gözlemlerin birbirinden ayırt edilmesi için uzaklık ve benzerlik ölçümleri kullanılır ve bu sayede gözlemler gruplara ayrılır (Yaz, 2014).

Uzaklık fonksiyonunun genel özellikleri ise şunlardır (Dinler, 2014):

- $d(j,i) = d(i,j)$; Uzaklık fonksiyonları simetriktir

- $d(j,j)=0$; Her birimin kendisine olan uzaklığı sıfırdır
- $d(i,j) \geq 0$; Uzaklık asla negatif olamaz
- $d(j,i) \leq d(j,h) + d(h,i)$; İki birimin üçüncü bir birime olan uzaklıkları toplamı bu iki birimin arasındaki uzaklıktan büyük olamaz (üçgen eşitsizliği).

Öklid, Karesel Öklid, Manhattan, Minkowski, Mahalanobis ve Chebyshev uzaklıkları en çok tercih edilen uzaklık ölçüleridir.

3.1.2.1 Öklid Uzaklığı

Öklid uzaklığı kümeleme analizinin en çok tercih edilen uzaklık ölçüsüdür. Genel olarak çok boyutlu uzayda geometrik uzaklıktır. Eşitlik (3.1)'de ki gibi hesaplanmaktadır (Tatlıdil, 1996).

$$d(x_i - y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.1)$$

Öklid uzaklığı formülleri işlenmiş veriler kullanarak hesaplama yapar, standartlaştırılmış veriler kullanmaz. Öklid uzaklıklarını önemli ölçüde boyutlar arasındaki ölçek farklılıkları etkilemektedir. Ancak kümeleme analizine sıra dışı olabilecek yeni nesnelerin eklenmesinden Öklid uzaklıkları etkilenmezler (Demiralay, 2005).

3.1.2.2 Karesel Öklid Uzaklığı:

Uzak nesnelerin daha büyük ağırlıklara sahip olabilmesi için Öklid uzaklığının karesi alınabilir. Karesel Öklid uzaklığı, aşağıdaki bağıntıda görüldüğü gibi, Öklid uzaklığı formülünün karesi alınarak hesaplanır.

$$d(x_i - y_i) = \sum_{i=1}^n (x_i - y_i)^2 \quad (3.2)$$

Burada; x_i ve y_i , gözlem çifti, n , nicel değişken sayısı ve $d(x_i - y_i)$, gözlem çifti arasındaki uzaklıktır.

3.1.2.3 Minkowski Uzaklığı:

Minkowski uzaklığı, metrik bir uzaklık ölçüsüdür ve nicel verilerin benzerliğinin hesaplanmasında kullanılır. Veri seti içerisindeki birbirlerinden iyi ayrılmış ve kompakt kümeleri belirlemede Minkowski uzaklığı iyi sonuçlar vermektedir (Mao ve Jain, 1996). Aşağıdaki gibi hesaplanmaktadır.

$$d(x_i - y_i) = \left(\sum_{i=1}^n \binom{n}{i} |x_i - y_i|^k \right)^{1/k} \quad (3.3)$$

3.1.2.4 Manhattan (City-Block) Uzaklığı:

Manhattan uzaklık ölçüsünde hesaplama yapılırken birimler arasındaki uzaklığın mutlak değeri alınır. Birimlerin aynı değişkenleri arasındaki mutlak farkların toplanması Formül (3.4) ile hesaplanır.

$$d(x_i - y_i) = \sum_{i=1}^n \binom{n}{i} |x_i - y_i| \quad (3.4)$$

Manhattan uzaklık ölçüsü uygulamada bazı problemlere yol açmaktadır. Manhattan uzaklık ölçüsünün değişkenler arasında herhangi bir ilişki olmadığını varsayarak işlem yapması bu problemlerden en göze çarpanıdır. Eğer araştırma konusunda aralarında korelasyon olan değişkenler varsa manhattan uzaklık ölçüsü kullanılarak hesaplanan uzaklık ölçüleri ile yapılan kümeleme bir anlam ifade etmeyecektir. Manhattan uzaklık ölçüsünün bir diğer problemi de eğer kullanılan değişkenlerinin birimleri aynı değilse bu uzaklık ölçüsünün anlamlı sonuçlar oluşturmadığı standartlaştırılmış karesel öklid uzaklığıyla karşılaştırıldığında görülebilmektedir (Atbaş, 2008).

3.1.2.5 Mahalanobis Uzaklığı:

Mahalanobis Uzaklık ölçüsü doğrudan birleştirme yapan, standart bir yöntemdir. Değişkenler arasındaki korelasyon veya kovaryansı dikkate aldığı için Mahalanobis uzaklığının, iki değişken arasında bir ilişki varsa kullanılması gerekmektedir (Dinler, 2014).

$$d(x_i - x_j) = D^2 = (x_i - x_j)' S^{-1} (x_i - x_j) \quad (3.5)$$

$p \times p$ kovaryans matrisini S göstermektedir. Aykırı noktaları da hesaplaması Mahalanobis uzaklığının en büyük avantajıdır (Sharma 1996).

3.1.2.6 Chebyshev Uzaklığı:

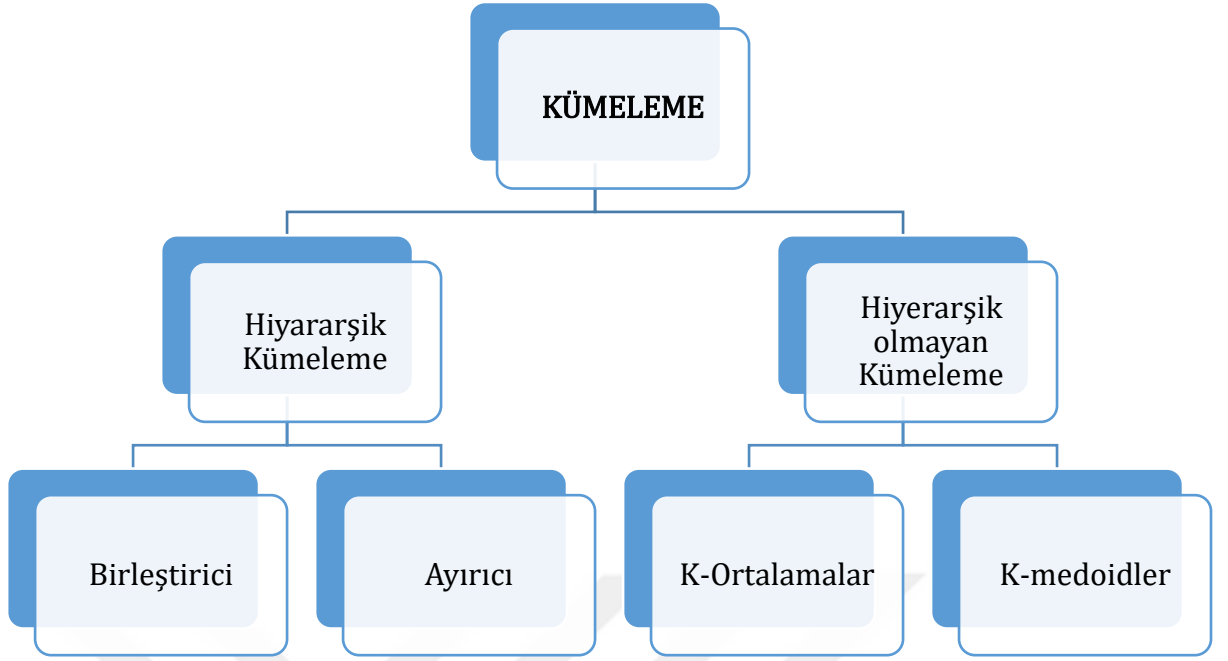
Chebyshev uzaklığı, iki gözlem vektörü arasındaki en yüksek mesafeyi ifade etmektedir. Terim büyüklüğünün farklı olduğu veri setlerinde, uzaklığın hesaplanmasında terim büyüklüğünün fazla olması dışındaki diğer boyutlarda değişik özellikler göz önünde bulundurulmayacağından bu uzaklığın kullanılması önerilmemektedir (Karaoğlu, 2018).

$$d_{ij} = \max |x_{ik} - x_{jk}| \quad (3.6)$$

3.1.3 Kümeleme Analizi Yöntemleri

Analizi gerçekleştirmek için sadece hangi uzaklık ölçüsünün kullanılacağını belirlemek yetmez, araştırmacı kümeleme işleminin nasıl yapılacağını da belirlemek zorundadır. Çeşitli yaklaşımlar kullanılarak birimlerin benzerliklerine göre kümelere ayrılması işlemi gerçekleştirilebilir. Bu yaklaşımlardan biri, birbirine en çok benzeyen iki veriyi aynı gruba atarak başlayıp daha sonra benzerlik seviyelerine göre verileri adım adım aynı kümede toplayarak devam eden ve tüm verilerin aynı kümeye atanması ile son bulan hiyerarşik bir yaklaşımdır. Bir diğer yaklaşım ise bütün verilerin hesaplanan ortalama değerlerine en yakın değerlere sahip olan verilerin aynı kümede toplanmasını sağlayan yaklaşımdır. Hiyerarşik ve hiyerarşik olmayan kümeleme adı altında en fazla kullanılan kümeleme yöntemleri iki gruba ayrılmaktadır (Blashfield ve Aldenderfer 1978).

Bir araştırmada hem hiyerarşik hem de hiyerarşik olmayan kümeleme yöntemlerinin ikisinin de beraber kullanılmasında fayda vardır. Çünkü hem çıkan sonuçlar hem de sonuçlara göre hangi yöntemin daha uygun olduğu karşılaştırılabilir.



Tablo 3.1 Kümeleme yöntemleri

3.1.3.1 Hiyerarşik Kümeleme Yöntemi

Veri setindeki birimlerin birbirlerine olan uzaklıklarını kullanarak, onların hiyerarşik ayrıştırmasını yapan yöntem hiyerarşik kümeleme yöntemleri denir. Birleştirici ve ayırıcı olmak üzere iki yöntem mevcuttur (Dinler, 2014).

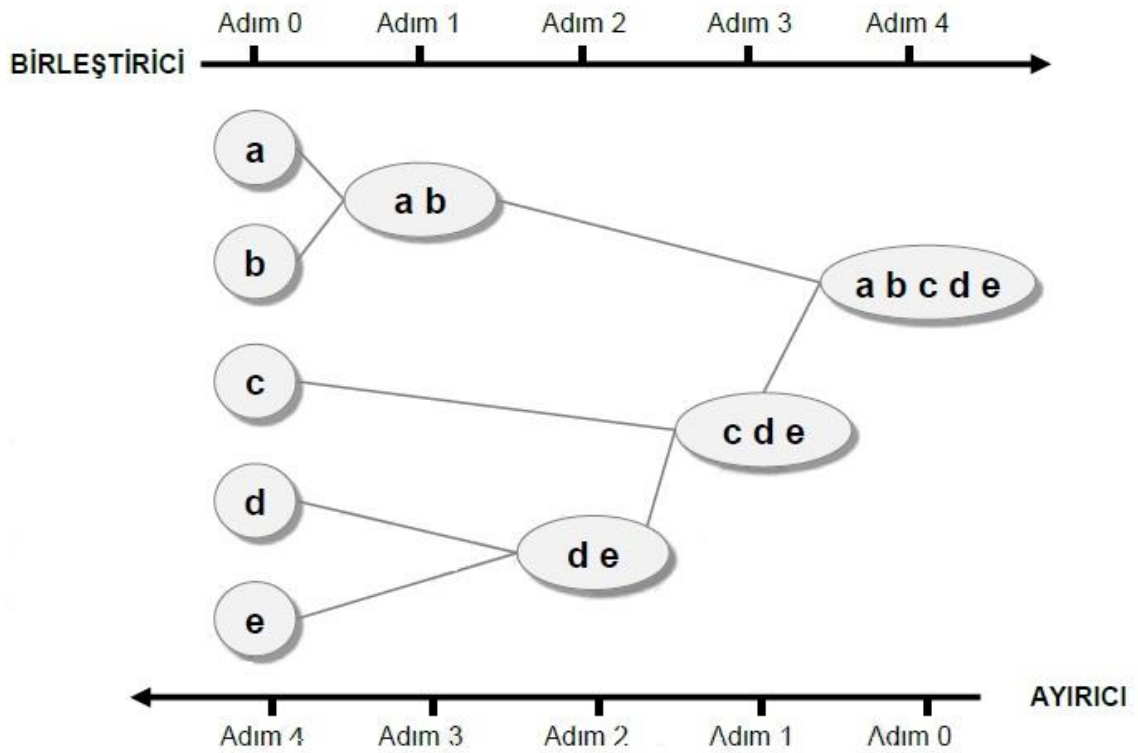
Hiyerarşik kümelemenin en yaygın kullanılan yöntemlerinden biri olan birleştirici yöntemde ilk olarak her bir veri tek başına ayrı bir küme olarak kabul edilmektedir. Belirli adımlardan sonra sürecin sonunda tüm veriler veya nesneler tek bir kümede toplanır.

Bu sürecin işleyişi şu şekilde özetlenebilir (Tatlıdil, 1996).

1. Öncelikle n adet bireyin her biri ayrı bir küme olmak üzere işleme başlanır
2. $d(i,j)$ değeri en küçük olan (en yakın iki küme) birleştirilir
3. Küme sayısı bir indirgenerek yinelenmiş uzaklıklar matrisi bulunur
4. 2 ve 3 numaralı adımlar $n-1$ kez tekrarlanır

Ayırıcı hiyerarşik yöntem de ise takip edilen süreç birleştirici hiyerarşik yöntemin tam tersi şeklindedir. Bu yöntemde bütün verilerden oluşan büyük tek bir küme ile işleme başlanır. Benzer olmayan kümeler farklılık büyüklüklerine göre adım adım ayrılarak yeni kümeleri oluştururlar. Her veri tek başına bir küme olana kadar işleme devam edilir. Bu yöntem uygun küme sayısının ne olduğu konusunda bilgi vermez (Everitt, 1974).

Birleştirici ve ayırıcı hiyerarşik kümeleme algoritmalarına ilişkin ağaç diyagramı şekil 3.2 de verilmiştir.

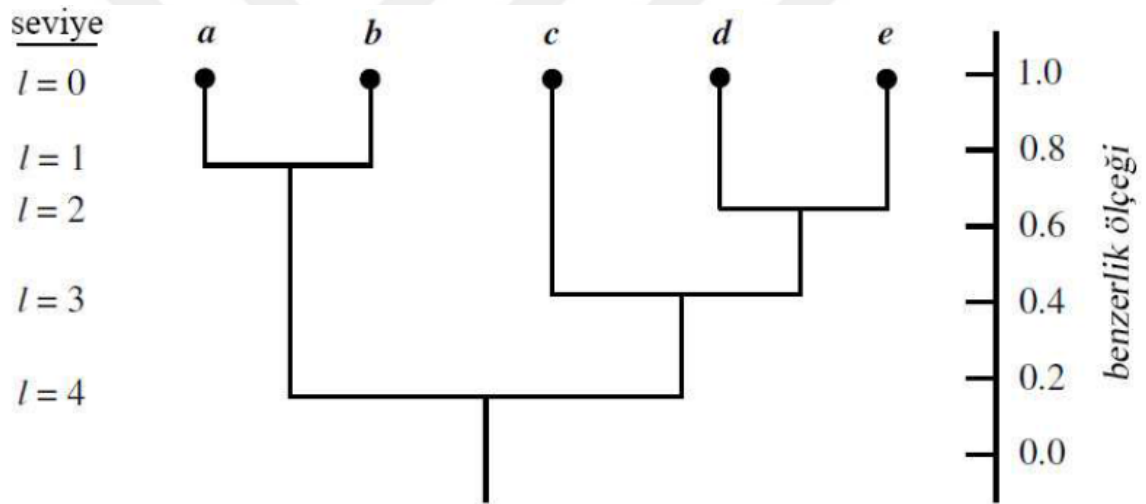


Şekil 3.2 Ayırıcı ve Birleştirici Hiyerarşik Kümelemenin {a,b,c,d,e} veri nesneleri üzerinde gösterimi (Han veKamber, 2006)

Şekil 3.2’de 5 adet nesnesi olan bir veri setinde {a,b,c,d,e} birleştirici ve ayırıcı hiyerarşik kümeleme algoritmalarının uygulamaları gösterilmektedir. Birleştirici yöntem başlangıçta her nesneyi birbirinden bağımsız bir şekilde ayrı küme olarak kabul eder. Daha sonra bazı kriterler doğrultusunda kümeler adım adım birleştirilir. Ayırıcı yöntemde ise başlangıçta bütün nesneler tek bir küme olarak kabul edilir. Daha sonra küme, maksimum öklid uzaklığı gibi bazı ölçütlere göre küme içerisinde yakın komşuluk ilişkisi olan objeler arasından kümelere ayrılır.

Küme bölme işlemi her nesne tek başına bir küme oluşturana kadar devam eder (Han ve Kamber, 2006).

Dendrogram olarak adlandırılan ağaç yapı nesnelerin adım adım nasıl gruplandığını (hiyerarşik kümeleme sürecini) göstermek için kullanılır. 5 nesne için bir dendrogram Şekil 3.3'te gösterilmiştir. $l=0$ seviyesinde her nesne tek başına bir küme ifade etmektedir. $l=1$ seviyesinde c, d ve e nesnelerinin her biri tek başına ayrı bir küme ifade ederken, ilk kümeyi a ve b nesneleri birleşerek oluşturmuştur. $l=2$ seviyesinde ise d ve e nesneleri birleşerek yeni bir küme oluşturmuşlardır. Ayrıca dikey eksen kümeler arasındaki benzerlik ölçeğini göstermek üzere kullanılabilir. Örneğin; $l=4$ seviyesinde bir araya gelerek tek bir küme oluşturan {c,d,e} ve {a,b} nesne grupları arasında yaklaşık 0.16 benzerlik vardır (Han ve Kamber, 2006).



Şekil 3.3 {a,b,c,d,e} Veri Nesnelerinin Hiyerarşik Kümeleneş Sürecini Gösteren Dendrogram Gösterimi (Han ve Kamber, 2006)

Hiyerarşik kümelemenin avantajlarını aşağıdaki gibi sıralayabiliriz.

- Dğümlerin seviyesine ilişkin esneklik özelliğinin olması
- Uzaklık ve benzerlik biçimlerini ele almanın kolaylığı
- Çeşitli niteliklere uygulanabilme özelliğinin olması

Hiyerarşik kümelemenin dezavantajlarını ise şunlardır:

- Bitirme kriteri belirsizdir ve
- Arada bulunan kümelere tekrar ulaşılmasına izin vermez (Güler 2006).

En fazla kullanılan 7 hiyerarşik kümeleme yöntem:

1. Tek Bağlantı Kümeleme Yöntemi (Single Linkage)
2. Tam Bağlantı Kümeleme Yöntemi (Complete Linkage)
3. Ortalama Bağlantı Kümeleme Yöntemi (Average Linkage)
4. McQuitty Bağlantı Kümeleme Yöntemi (McQuitty Linkage)
5. Küresel (Merkezi) Bağlantı Kümeleme Yöntemi (Centroid Linkage)
6. Medyan Bağlantı Kümeleme Yöntemi (Median Linkage)
7. Ward Bağlantı Kümeleme Yöntemi (Ward Linkage)

Tek Bağlantı Kümeleme Yöntemi (Single Linkage):

İlk olarak Florek vd. (1951) tarafından uygulanan bu teknik daha sonra sırasıyla Sneath (1957) ve Johnson (1967) tarafından uygulanmıştır. Birbirine en yakın iki gözlem veya kümeyi birleştirmek için uzaklık veya benzerlik matrislerinden yararlanılır ve yaptığı birleştirme işlemi yinelenmektedir (Şentürk 1995, Fırat 1997).

En yalın hiyerarşik kümeleme yöntemi olan bu yöntem aynı zamanda en yakın komşuluk yöntemi olarak da bilinmektedir. Bu yöntemde her birim herhangi bir kümeye atanana kadar uzaklık matrisinden faydalanılarak birbirine en yakın kümeler veya birimler birleştirilir. Bir birim tek başına bir küme de oluşturabilir yani birleştirme yapılırken kümelerin eleman sayısının birden fazla olması koşulu yoktur. Bir kümede ki bir gözlem ile diğer kümedeki bir gözlem arasındaki minimum mesafe iki küme arasındaki uzaklıktır (Dinler, 2014).

Birbirinden çok az farklı olan kümeleri ayırmada tek bağlantı yöntemi yetersiz kalmaktadır. Yöntem, elips şeklinde dağılmayan değerleri aynı kümede toplayabilen az sayıdaki kümeleme yönteminden biridir. Bu nedenle yöntemin uygulanması sonucunda çoğu kez birbirinden oldukça farklı birimlerin aynı kümede yer alması mümkün olabilmektedir (Anderberg, 1973).

Tam Bağlantı Kümeleme Yöntemi (Complete Linkage):

En uzak komşuluk yöntemi olarak da bilinen tam bağlantılı kümeleme yöntemi, tek bağlantı kümeleme yöntemine oldukça benzeyen bir diğer kümeleme yöntemidir. Tek bağlantı tekniğinden farklı olarak, birimler arası minimum uzaklık yerine

maksimum uzaklıklar hesaplanır. Aralarındaki uzaklıkları en az olan birimlerin aynı kümede toplanmasıyla, tek bağlantı tekniğindeki gibi kümeleme işlemi gerçekleştirilir (Tatlidil, 1996).

Küme içindeki bireylerin uzaklıklarının belirli bir değerden küçük olması durumunda bütün kümelerin iyi sonuç vermesini garanti etmemektedir (Malkoç, 2018).

Ortalama Bağlantı Kümeleme Yöntemi (Average Linkage):

Bir küme arasındaki eleman çiftleri ile diğer bir kümedeki eleman çiftleri arasındaki ortalama fark alınarak iki küme arasındaki fark bulunur. Bu yöntem Sokal ve Michener tarafından önerilmiştir (Everitt 1981). Başlangıç olarak aşırı uç gözlemlere bakılmaz. Kümenin ortasına düşen gözlemi esas alır (Kalaycı, 2010).

Tek bağlantı ve tam bağlantı yöntemleri arasında sonuç vermesinden dolayı ortalama bağlantı yönteminden alternatif bir yöntem olarak bahsedilmektedir (Hubert ve Levin, 1976).

McQuitty Bağlantı Kümeleme Yöntemi (McQuitty Linkage):

Ağırlıksız ortalama bağlantı kümeleme yöntemi olarak da bilinen yöntemde iki kümenin birleştirilmesi esnasında bir kümenin diğer kümeye mesafesi yakında birleştirilecek kümelerin mesafelerinin ortalaması olarak hesaplanır (Yazgan ve Kayaalp 2002).

Küresel (Merkezi) Bağlantı Kümeleme Yöntemi (Centroid Linkage):

Ortalama bağlantı kümeleme yönteminin özel bir biçimi olan bu yöntemde kümeler arası uzaklıklar ve küme merkezleri arası uzaklıklar olarak tanımlanır. Küme merkezlerinin birbirlerine olan uzaklıklarına göre kümelerin birleştirilmesi işlemi yapıldığı için kümeler merkezleri ile ifade edilir (Renchber, 2002).

İki küme arasındaki uzaklık küme merkezleri arasındaki Karesel Öklid uzaklığıdır. Merkez değeri, kümeyi oluşturan gözlemlerin tüm değişkenlere göre ortalamasını göstermektedir. Farklı nitelikteki gözlemlerden çok fazla etkilenmemesi bu yöntemin en önemli avantajıdır (Lattin vd., 2003).

Medyan Bağlantı Kümeleme Yöntemi (Median Linkage):

Merkezi kümeleme algoritması metodunda birleştirilecek olan iki grubun boyutları arasında oldukça büyük fark varsa, bu iki grubun birleşmesi sonucunda oluşacak olan yeni kümenin merkezi, boyutu büyük olan kümeye daha yakın olacaktır (Everitt, 1993).

Medyan kümeleme algoritmasıyla oluşturulan yeni grubun merkezi grupların boyutundan bağımsızdır. Ancak, her algoritmada olduğu gibi bu algoritmanın da dezavantajı bulunmaktadır. Geometrik olarak yorumlanamaması bu yöntemin dezavantajıdır. Bundan dolayı da, kolerasyon katsayıları gibi ölçüler bu yöntem için uygun değildir (Gan vd. 2007).

Ward Bağlantı Kümeleme Yöntemi (Ward Linkage):

Minimum varyans tekniği olarak da isimlendirilen Ward Tekniği, Joe H. Ward tarafından 1963 yılında ortaya atılmış olup bu yöntemde kümeler arasındaki uzaklıkları hesaplamak yerine, küme içindeki homojenliğin maksimizasyonuna odaklanılmaktadır. Homojenliğin ölçüsü olarak da küme içi kareler toplamı kullanılmakta ve bunun minimize edilmesine çalışılmaktadır. Minimum değeri araştırılan küme içi kareler toplamı aynı zamanda Hata Kareler Toplamı (HKT) olarak da adlandırılmaktadır (Sharma, 1996).

Ward tekniği, varyansları baz alarak birleştirme işlemini yapar ve birleştirme işlemine değişkenliği en az olan kümeler ile başlar. Küme içi kareler toplamının en küçük olduğu durumda en iyi sonucu vermektedir (Koldere 2008). İki küme birleştirildiğinde grup içi hata diğer bir kümeyle birleştirilmesine kıyasla daha az olacaksa uygulama sonlanır, aksi bir durum söz konusuysa üçüncü olan diğer küme ile birleşim sağlanır (Hair vd., 2006).

Ward metodunun sonuçları genellikle optimal değerler vermektedir. Kuiper and Fisher 1975 yılında Monte Carlo metodu kullanarak diğer hiyerarşik kümeleme algoritmalarıyla Ward metodunun karşılaştırmasını sunmuştur (Gan vd. 2007).

Hiyerarşik kümeleme algoritmaları, pek çok sayıda benzerlik veya uzaklık ölçüsü bulunduğundan dolayı ve bunların seçimlerinin araştırmacının amaçlarına göre

değişiklik göstermesinden kaynaklı esnek bir yapıya sahiptir. Bu durum hiyerarşik kümeleme tekniklerinin olumlu yönlerinden sadece biridir. Ayrıca, hiyerarşik olmayan kümeleme tekniklerindeki gibi hiyerarşik kümeleme tekniklerinde kümeleme yapılmadan önce küme sayısının belirlenmesine gerek yoktur. Bu yönüyle de hiyerarşik kümeleme teknikleri avantajlıdır. Ancak, herhangi bir kümeleme tekniğinde olduğu gibi bu hiyerarşik kümeleme tekniklerinde de belli başlı dezavantajlar bulunmaktadır. Örneğin, uygulamada kullanılacak olan veri tabanı büyük olsun. Bu durumda, oluşacak olan benzersizlik matrisi de oldukça büyük olacağından dolayı, çoğu istatistiksel bilgisayar yazılımlarından sonuç elde edilemeyebilir. Buna ek olarak, yöntem aykırı değerlerden kolaylıkla etkilenmektedir (Fırat, 2016).

3.1.3.2 Hiyerarşik Olmayan Kümeleme Yöntemi

En temel kümeleme analizi tekniklerinden biri olan hiyerarşik olmayan kümeleme tekniklerinde, n ile veri tabanındaki nesne sayısı ve k ile oluşturulacak küme sayısı gösterilsin. $k \leq n$ koşulunu sağlamak şartıyla hiyerarşik olmayan kümeleme teknikleri n adet nesneyi k adet kümeye bölmeler. Böylece, oluşan her bölüm bir kümeyi gösterir. Kümeler tarafsız bölme ölçütü olarak nitelendirilen bir ölçüte uygun oluşturulduğu için aynı kümede yer alan nesneler birbirlerine benzerken, farklı kümede bulunan nesneler birbirinden farklıdır (Kaya ve Köymen, 2008).

Elde edilmesi amaçlanan küme sayısı bazı durumlarda önceden bellidir ve bu küme sayısına sadık kalınarak çözümler üretilmesi gerekmektedir. Hiyerarşik olmayan teknikler, küme sayısı ile alakalı herhangi bir bilgi varsa veya küme sayısına karar verilmiş ise bu tür durumlarda uygulanması için genellikle uzun zamana ihtiyaç duyulan hiyerarşik teknikler yerine tercih edilmektedir (Anderberg, 1973).

Birim sayısının fazla olması durumunda, adım adım birimleri birbirine bağlayarak kümeleri oluşturmak oldukça zaman alan bir işlemdir. Kümelenmesi gereken çok sayıda birimin bulunduğu uygulamalarda, ağır bir iş yükü gerektiren hiyerarşik algoritmalar yerine hiyerarşik (aşamalı) olmayan kümeleme yöntemleri kullanılmaktadır (Erişoğlu 2011). Hiyerarşik olmayan kümeleme yöntemi;

birimlerin birbirleri ile olan ortak özelliklerini dikkate alarak bu özelliklere benzeyen vektörleri tanımlayan, kendi içinde benzerliğin maksimum ve aralarında ise benzerliğin minimum olduğu gruplar oluşturmayı amaçlamaktadır (Hartigan, 1975).

Hiyerarşik olmayan kümeleme yöntemleri değişkenleri, gruplandırmaktan ziyade nesneleri gruplandırmaya yarar ve bu nesneleri K adet kümeye ayırır. Küme sayısı k başlangıçta belirlenebileceği gibi kümeleme sürecine bağlı olarak da ortaya çıkabilir. Hiyerarşik olmayan yöntemler çok daha büyük ve kompleks veri setlerine hiyerarşik olan yöntemlere göre çok daha rahat uygulanır (Johnson, 2002).

Hiyerarşik olmayan kümeleme yöntemlerinden en fazla kullanılanları k -ortalamalar ve k -medoids yöntemidir.

K-Ortalama Yöntemi:

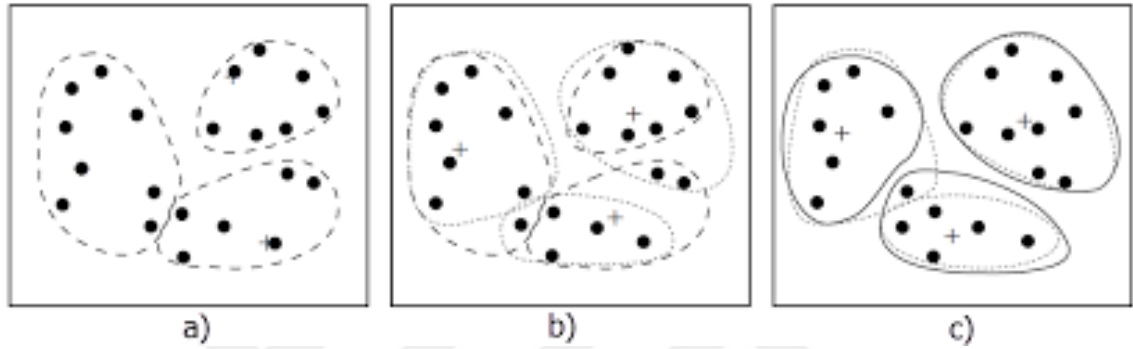
K -ortalamalar algoritması 1967 yılında Mac Quenn tarafından geliştirilen en eski kümeleme yöntemlerinden biridir. Bu algoritma, n adet veriden oluşan bir veri tabanında K adet küme oluşturur (Vattani, 2011).

Amaç bütün kümeleme yöntemlerinde olduğu gibi, gerçekleştirilen kümeleme işlemi sonucunda oluşacak kümelerin, küme içerisindeki benzerliklerin en üst seviyede, kümeler arası benzerliklerinin ise en düşük seviyede olmasını sağlamaktır. Kümedeki her birimin, kümenin ağırlık merkezi olarak kabul edilen bir birimine olan uzaklıklarının ortalama değeri alınarak küme benzerlikleri ölçülmektedir (Han and Kamber 2006). K -Ortalamalar yöntemindeki işlem adımları aşağıdaki gibidir (Dinler, 2014).

1. Başlangıç küme merkezleri olarak k adet birim rastgele seçilir.
2. Başlangıç küme merkezi olarak belirlenmemiş olan her birim belirlenen uzaklık ölçütlerine göre başlangıç küme merkezlerinin ait oldukları kümelere atanır.
3. Oluşturulan k adet başlangıç kümesindeki değişkenlerin ortalamaları alınarak yeni küme merkezleri oluşturulur.
4. Birimler yeni oluşturulan kümelerin küme merkezleri içerisinde en yakın mesafede oldukları kümeye atanırlar.

5. Yeni oluşturulan küme merkezlerine olan uzaklıklar ile bir önceki küme merkezlerine olan uzaklıklar karşılaştırılır.
6. Uzaklıklar kabul edilebilir oranda azalmış ise 4. Adıma dönülür.
7. Eğer çok büyük bir değişiklik meydana gelmemiş ise, tekrarlama işlemi sona erdirilir.

Şekil 3.4’de k-Ortalama metodu ile bir nesne setinin kümelenmesi gösterilmiştir. “+” ile her bir kümenin orta değeri işaretlenmiştir (Han ve Kamber, 2006).



Şekil 3.4 K-ortalama yöntemi ile kümeleme (Han ve Kamber, 2006)

K-Medoidler Yöntemi:

Hiyerarşik olmayan kümeleme yöntemleri içinde K-ortalamalar yöntemi gibi önemli bir yere sahip olan başka bir yöntem ise K-medoidler yöntemidir. 1990 yılında Kaufmann ve Rousseeuw tarafından bu yöntem geliştirilmiştir.

Temsilci (medoid) seçimi demek küme merkezine en yakın mesafede bulunan noktanın belirlenmesidir. k adet temsilcinin k adet küme için belirlenmesinden sonra, temsilci olmayan diğer noktaların her biri, kendilerine en fazla benzeyen temsilcinin etrafında toplanırlar (Silahtaroglu, 2013).

K-Medoids algoritmasında yapılacak işlem adımları aşağıdaki gibidir (Akın,2008):

1. Adım: k küme sayısının tayin edilmesi.
2. Adım: k nesnelerinin başlangıç medoidleri olarak seçimi.
3. Adım: Kalan nesnelerin en yakın medoid x' e sahip kümeye, atanması

4. Adım: Amaç fonksiyonunun hesaplanması. (Hata kareler kriteri: En yakın medoidler için bütün nesnelerin uzaklıklarının toplamı)
5. Adım: Tesadüfî olarak medoid olmayan y noktasının seçimi.
6. Adım: x ile y 'nin yer değiştirmesi sonucunda amaç fonksiyonu minimize olacaksa bu iki noktanın (x ile y) yerinin değiştirilmesi.
7. Adım: 3. Adım ile 6. Adım arası işlemlerin değişiklik olmayana kadar tekrarlanması.

3.1.3.3 Yöntemler Arasındaki Karşılaştırma

Bahsedilen kümeleme yöntemlerinden hangisi kullanılmalıdır sorusunun net bir cevabı bulunmamaktadır. Genel olarak hiyerarşik yöntemlerden; Ward ve Ortalama Bağlantı yöntemi en çok kullanılan yöntemlerdendir. Birim sayısının fazla olmasından dolayı benzerlik matrisinin büyük boyutlara ulaşması, uzaklık ölçüsü farklılığı ve aykırı değerlerin etkisi gibi nedenlerle hiyerarşik olmayan yöntemler de sık sık kullanılmaktadır. Bazı araştırmalarda bu iki yönteminde birlikte uygulanması söz konusu olmaktadır (Burmaoğlu 2011: 369-382).

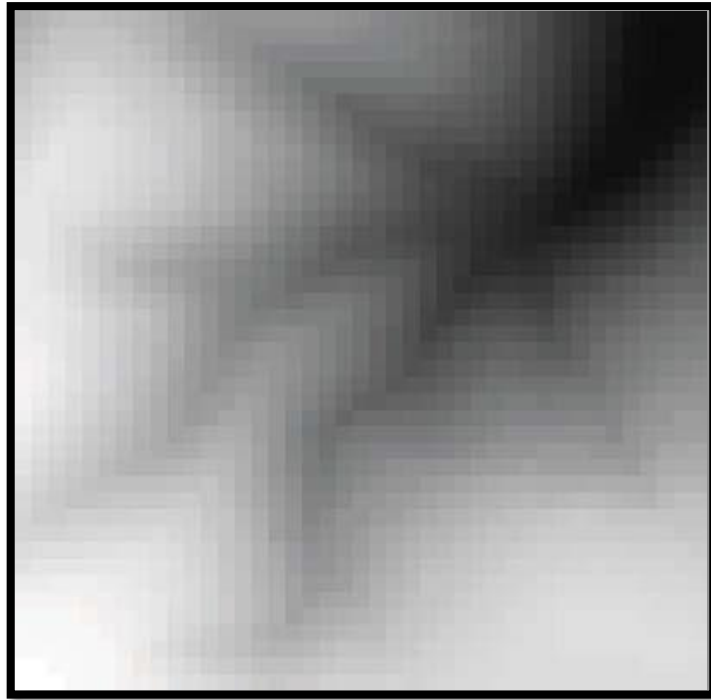
3.1.4 Benzerlik Ölçümleri

Kümeleme analizinin asıl amacı, gözlenen nesne ya da bireylerin arasındaki uzaklıkları ya da benzerlikleri belirlemektir. Benzerlik ve uzaklık kavramları birbirlerinin tersi olup, küçük bir sayı çıktığında iki nesnenin birbirinden uzak olduğunu, büyük bir sayı çıktığında ise iki nesnenin birbirine yakın olduğunu gösterir. Verilerin metrik ya da kategorik olmasına göre benzerlik ölçümü seçimi değişmektedir (Kalaycı, 2005).

Bu çalışmada iki kümeleme yöntemine (hiyerarşik ve K- ortalama) göre ikişer uygulama gerçekleştirilmiştir. Birincisinde kümeleme yönteminde kullanılan özniteliklerin kümeleme sonucuna etkilerinin eşit olması istenerek tüm özniteliklere eşit ağırlık verilmiştir. İkincisinde ise özniteliklerin önem derecesi göz önüne alınarak özniteliklere farklı ağırlıklar verilmiştir.

4.1 Veriler

Uygulamalarda Şekil 4.1’de görünen grid temelli SYM kullanılmıştır. Bu SYM 10 metre çözünürlüklü olup, 1:5,000 ölçekli standart topoğrafik harita verilerinden türetilmiştir ve 1681 adet grid noktasından (41 satır, 41 sütun) meydana gelmektedir. 160 adet grid noktası sınır boyunca yer almaktadır. Sınır noktaları hariç toplam da 1521 grid noktası bulunmaktadır. Kümeleme işlemine sınır noktaları dahil edilmemiştir.



Şekil 4.1 Uygulamada kullanılan SYM

4.2 Benzerlik Ölçütleri: Öznitelikler

Bu uygulamada grid noktalarının birbirlerine olan benzerliklerini belirlemek için aşağıdaki şu soruların cevapları aranmıştır.

- En düşük ya da en yüksek kotlu nokta mıdır?
- Grid noktasından geçen dört profil (Batı-Doğu, Güney-Kuzey, Güneybatı-Kuzeydoğu, Güneydoğu-Kuzeybatı) boyunca kendisinden bir önceki ve bir sonraki noktadan daha düşük ya da daha yüksek kotlu bir nokta (ekstrem nokta) mıdır?
- Sekiz komşusu (Doğu, Güneydoğu, Güney, Güneybatı, Batı, Kuzeybatı, Kuzey, Kuzeydoğu) arasında ekstrem nokta(lar) var mıdır? Varsa hangi komşusu yada komşuları ekstrem noktadır?

Böylece, her bir grid noktasının on üç özniteliği belirlenmiştir (Tablo 4.1).

Tablo 4.1 Grid noktalarının öznitelikleri (EY/ED: En Yüksek / En Düşük, D: Doğu, Gd: Güneydoğu, G: Güney, Gb: Güneybatı, B: Batı, Kb: Kuzeybatı, K: Kuzey, Kd: Kuzeydoğu)

Öznitelikler												
EY/ED	Ekstrem Noktalar				Komşu Noktalar							
	B-D	G-K	Gb-Kd	Gd-Kb	D	Gd	G	Gb	B	Kb	K	Kd

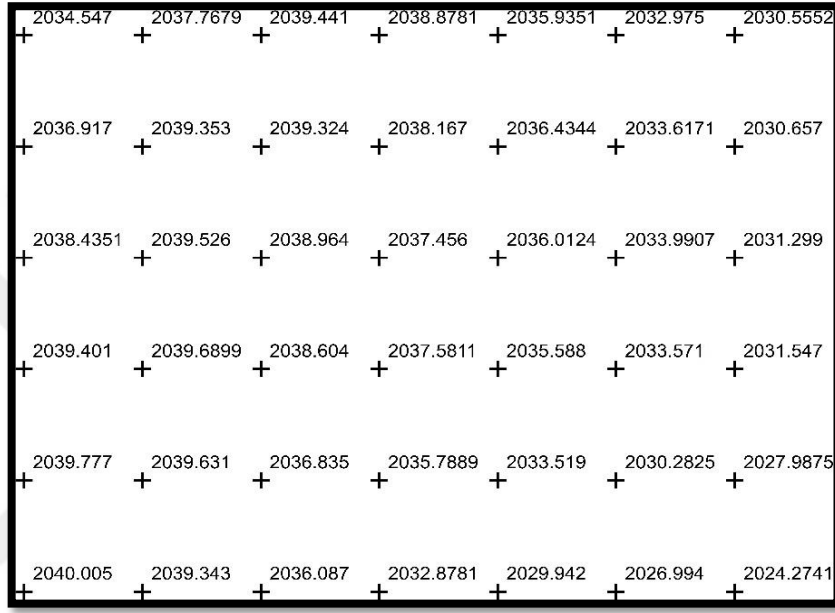
4.2.1 En Düşük ve En Yüksek Kotlu Nokta

Uygulamada kullanılan haritadaki sınır noktaları dışında kalan bütün noktaların kotlarına bakıldığı zaman 2050,6315m kotlu nokta en yüksek kotlu nokta, 1931,165m kotlu nokta ise en düşük kotlu nokta olarak belirlenmiştir.

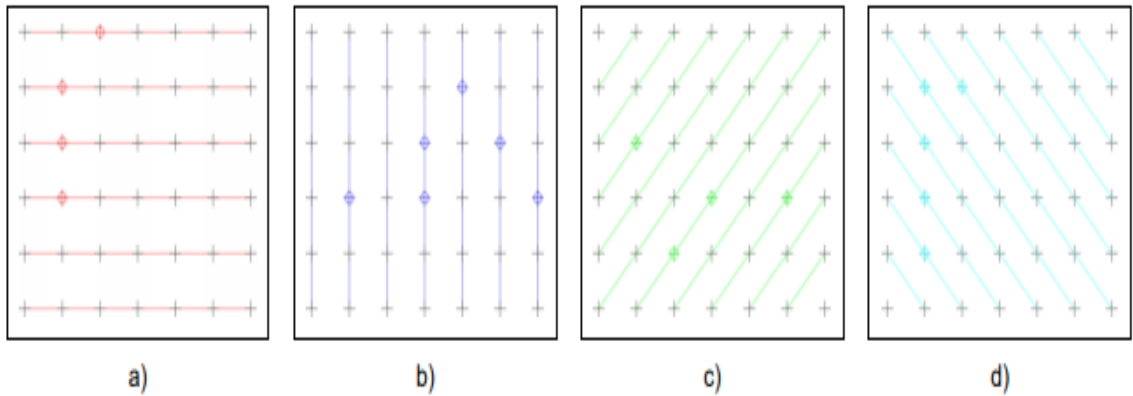
4.2.2 Ekstrem Noktalar

Bir grid noktasının, batı- doğu, güney-kuzey, güneybatı-kuzeydoğu ve güneydoğu-kuzeybatı profillerinde yer alan komşu noktalarının kotlarına bakılarak ekstrem nokta olup olmadığına karar verilir. Bir noktanın aynı profilde komşusu olan iki ardışık grid noktasının (örneğin batı ve doğusundaki grid noktaları) yükseklikleri o noktanın yüksekliğinden düşük yada fazla ise o nokta ekstrem noktadır.

Şekil 4.2 de, 42 adet grid noktasından meydana gelen bir SYM, her bir grid noktasının yükseklik değeri ile birlikte görünmektedir. Şekil 4.3a'da Batı-Doğu profilleri ve bu profiller boyunca belirlenen ekstrem noktalar, Şekil 4.3b'de GüneyKuzey profilleri ve bu profiller boyunca belirlenen ekstrem noktalar, Şekil 4.3c'de Güneybatı-Kuzeydoğu profilleri ve bu profiller boyunca belirlenen ekstrem noktalar ve Şekil 4.3d'de Güneydoğu-Kuzeybatı profilleri ve bu profiller boyunca belirlenen ekstrem noktalar görünmektedir.

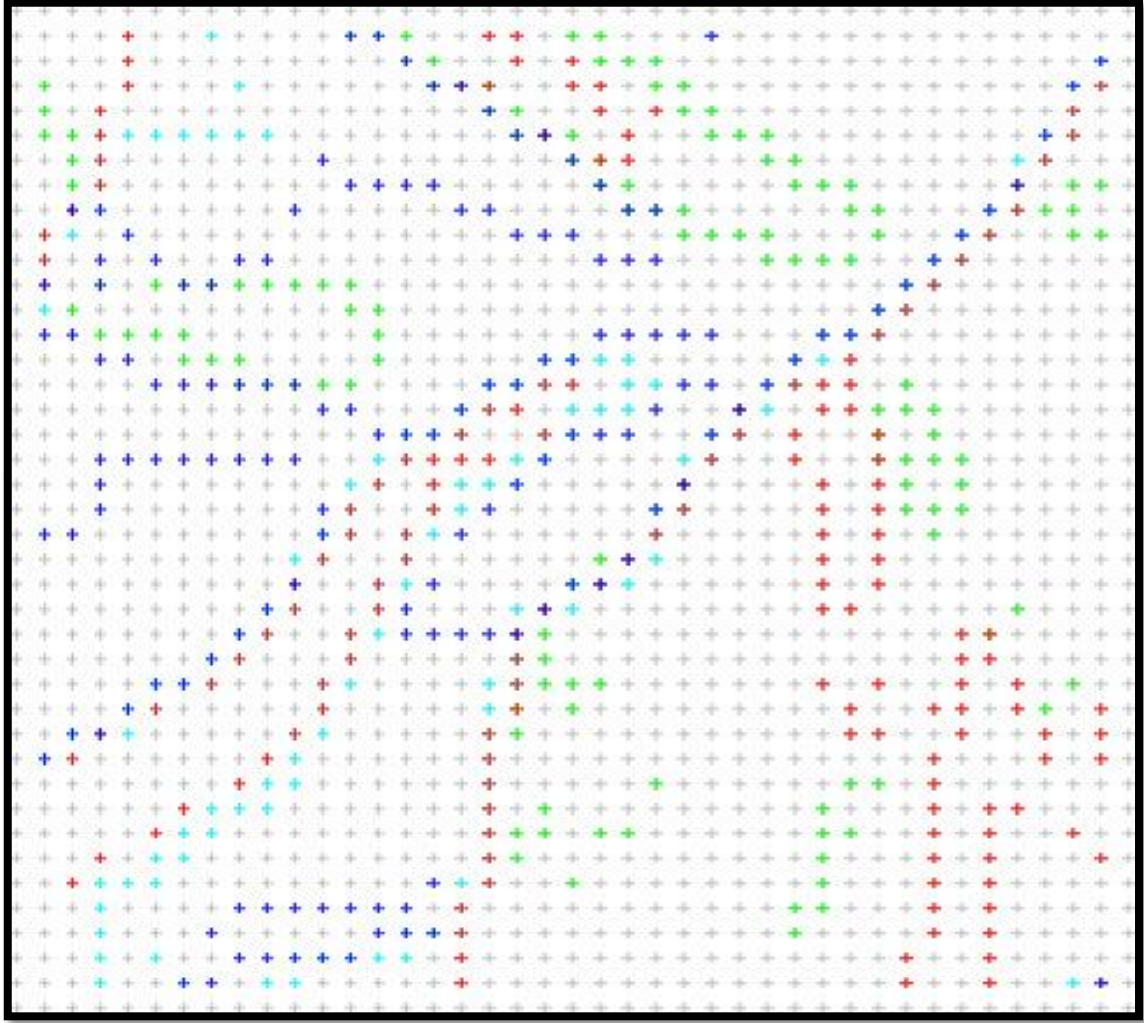


Şekil 4.2 Örnek grid noktaları ve yükseklik değerleri



Şekil 4.3 a) Batı-Doğu profilleri ve bu profiller boyunca belirlenen ekstrem noktalar, b) Güney-Kuzey profilleri ve bu profiller boyunca belirlenen ekstrem noktalar, c) Güneybatı-Kuzeydoğu profilleri ve bu profiller boyunca belirlenen ekstrem noktalar ve d) Güneydoğu-Kuzeybatı profilleri ve bu profiller boyunca belirlenen ekstrem noktalar

Uygulamada kullanılan SYM'deki 1521 noktadan 466'sının ekstrem nokta olduğu tespit edilmiştir (Şekil 4.4).



Şekil 4.4 Batı-Doğu (kırmızı), Güney-Kuzey (mavi), Güneybatı-Kuzeydoğu (yeşil) ve Güneydoğu-Kuzeybatı (gök mavisi) profilleri boyunca belirlenen ekstrem grid noktaları

4.2.3 Eşit Ağırlıklı Öznitelikler

Yukarıdaki sorular grid noktalarının her birine teker teker sorulduğunda cevabı evet olan özniteliklere 1 (bir), hayır olan özniteliklere ise 0 (sıfır) değeri atanmıştır (Tablo 4.2). Değerler bu şekilde verildiği için tüm özniteliklerin kümeleme analizinin sonucuna etkisi eşit miktarda olmuştur.

Tablo 4.2 Grid noktalarının eşit ağırlıklı öznitelikleri

Nokta No	Öznitelikler												
	EY/ED	Ekstrem Noktalar				Komşu Noktalar							
		B-D	G-K	Gb-Kd	Gd-Kb	D	Gd	G	Gb	B	Kb	K	Kd
1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0	0	0	0	1
3	0	0	0	0	1	0	0	0	0	0	0	1	0
4	0	0	0	0	0	0	0	0	0	1	1	0	0
5	0	0	0	0	0	1	0	0	0	0	0	1	0
6	0	0	1	0	0	1	0	0	0	0	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1521	0	0	0	0	0	1	0	1	0	0	0	0	0

4.2.4 Ağırlıklandırılmış Öznitelikler

Özniteliklerin önemine göre kümeleme sonucuna etki miktarının değiştirilmesi istenmiştir. Bu durumda grid noktası, en düşük veya en yüksek kotlu nokta ise ilgili öznitelik bölümüne 10 (on), kendisinin ekstrem nokta olması durumunda 2 (iki), komşularından biri ekstrem nokta ise 1 (bir) değeri atanmıştır (Tablo 4.3).

Tablo 4.3 Grid noktalarının ağırlıklandırılmış öznitelikleri

Nokta No	Öznitelikler												
	EY/ED	Ekstrem Noktalar				Komşu Noktalar							
		B-D	G-K	Gb-Kd	Gd-Kb	D	Gd	G	Gb	B	Kb	K	Kd
1	10	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0	0	0	0	1
3	0	0	0	0	2	0	0	0	0	0	0	1	0
4	0	0	0	0	0	0	0	0	0	1	1	0	0
5	0	0	0	0	0	1	0	0	0	0	0	1	0
6	0	0	2	0	0	1	0	0	0	0	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1521	0	0	0	0	0	1	0	1	0	0	0	0	0

4.3 Hiyerarşik Yöntemle Kümeleme

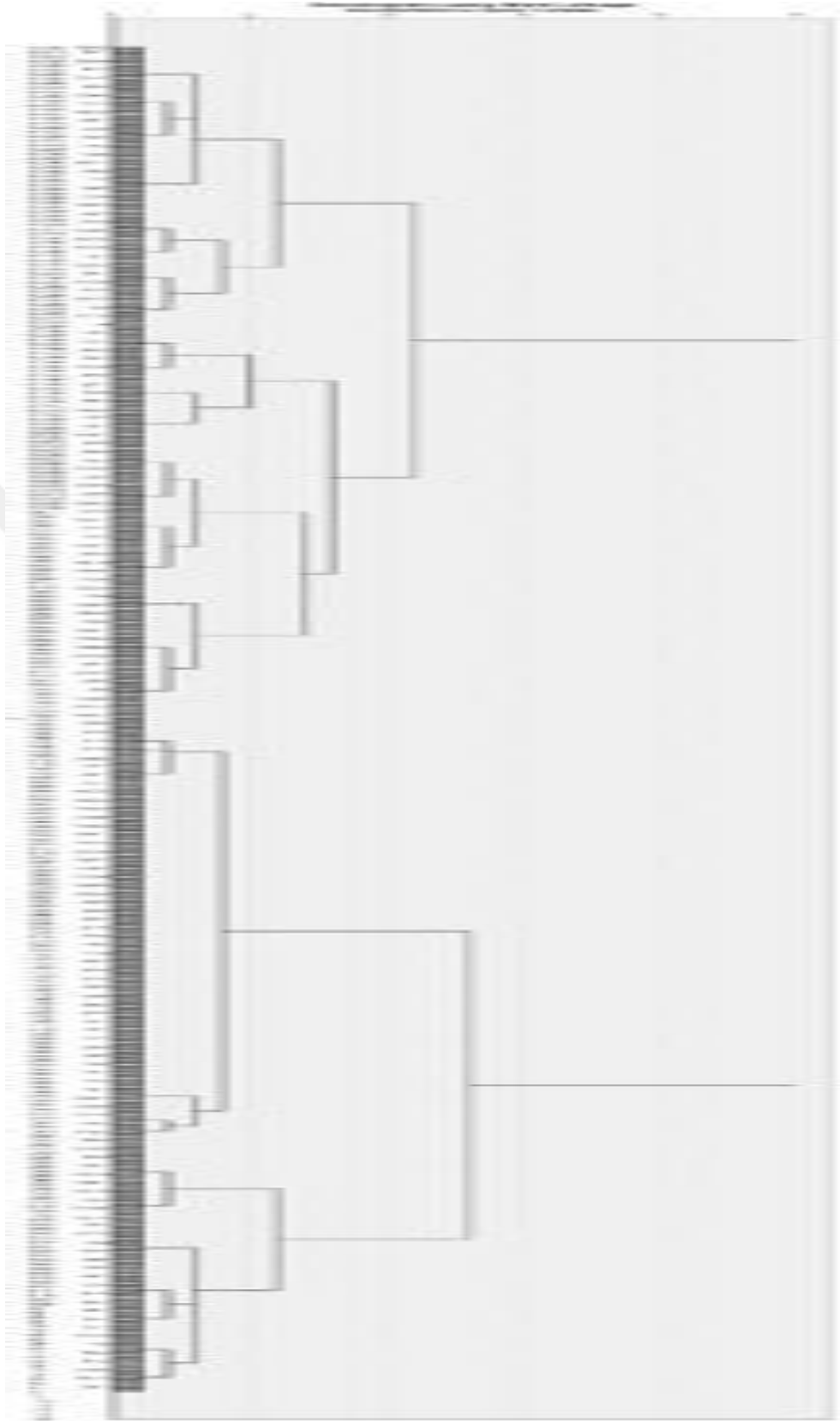
Hem eşit ağırlık öznitelikler kullanılarak hem de ağırlıklandırılmış öznitelikler ile yapılan kümeleme analizi işlemi SPSS uygulaması kullanılarak yapılmıştır. Her iki kümeleme işlemi de Ward yöntemine göre, yakınlık ölçütü olarak Öklid mesafesinin karesi kullanılarak yapılmıştır.

4.3.1 Eşit Ağırlıklı Öznitelikler İle Kümeleme

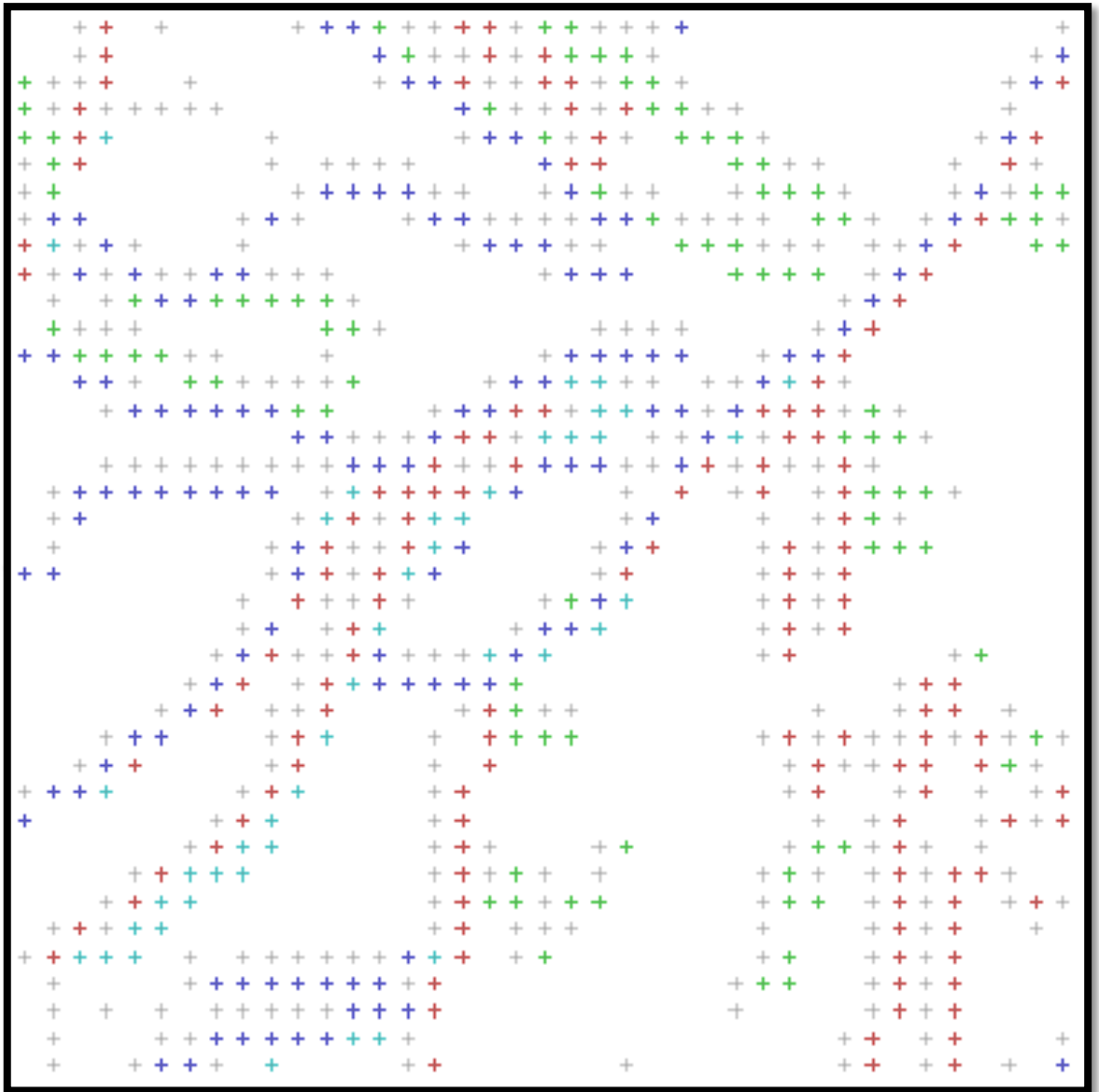
Küme sayısı 2 olarak belirlenmiştir. Kümeleme işlemi sonunda Şekil 4.5’te görünen dendogram oluşmuştur. Bu kümelerin içerikleri incelendiğinde,

- Birinci kümede, 415’i ekstrem nokta özelliğine sahip, en yüksek ve en düşük kotlu noktayı da içeren toplam 742 nokta ve
- İkinci kümede, 51’i ekstrem nokta özelliğine sahip toplam 779 nokta olduğu görülmüştür.

Sonuç olarak, ekstrem noktaların fazla olduğu ve aynı zamanda en yüksek ve en düşük kotlu noktaları da içeren birinci küme “Seçilenler”, diğeri ise “Elenenler” kümesi olarak kabul edilmiştir. Seçilenler kümesindeki noktalar Şekil 4.6’da görünmektedir



Şekil 4.5 Eşit ağırlıklı öznitelikler ile hiyerarşik kümeleme dendogramı



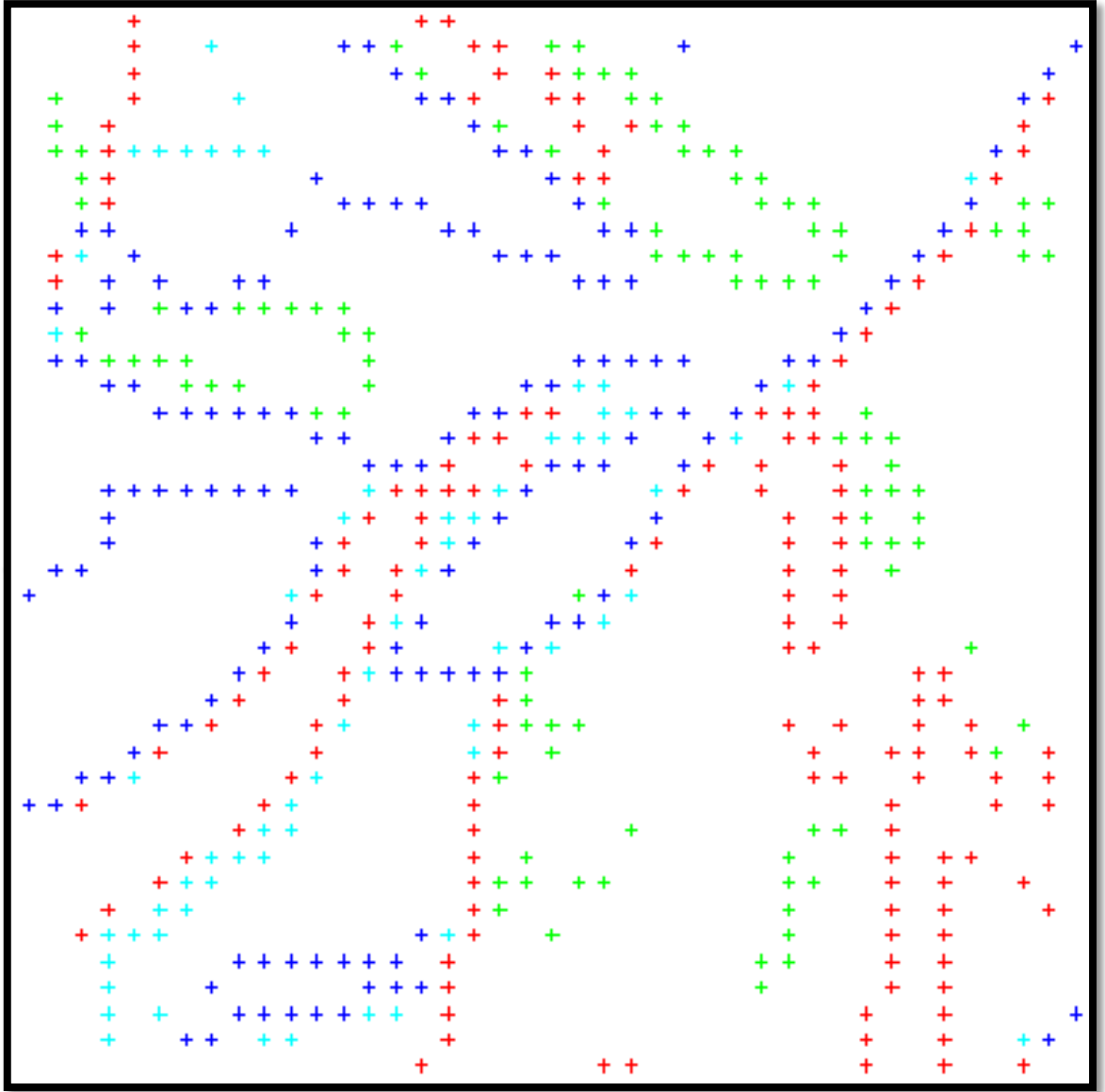
Şekil 4.6 Eşit ağırlıklı öznitelikler ile hiyerarşik kümeleme analizi sonucu seçilen, Batı-Doğu (kırmızı), Güney-Kuzey (mavi), Güneybatı-Kuzeydoğu (yeşil) ve Güneydoğu-Kuzeybatı (gök mavisi) profilleri boyunca belirlenen ekstrem grid noktaları, ekstrem nokta olmayıp seçilen grid noktaları (gri)

4.3.2 Ağırlıklandırılmış Öznitelikler İle Kümeleme

Küme sayısı 5 olarak belirtilmiştir. Kümeleme işlemi sonunda Şekil 4.7’de görünen dendogram oluşmuştur. Bu kümelerin içerikleri incelendiğinde,

- Birinci kümede, tamamı ekstrem nokta özelliğine sahip 146 nokta (Güney-Kuzey profilleri boyunca belirlenen ekstrem noktaların tamamı ve en yüksek ve en düşük kotlu noktalar),
- İkinci kümede, tamamı ekstrem nokta özelliğine sahip olmayan 1054 nokta,
- Üçüncü kümede, tamamı ekstrem nokta özelliğine sahip 62 nokta (Güneydoğu-Kuzeybatı profilleri boyunca belirlenen ekstrem noktaların tamamı),
- Dördüncü kümede, tamamı ekstrem nokta özelliğine sahip 149 nokta (Batı-Doğu profilleri boyunca belirlenen ekstrem noktaların tamamı) ve
- Beşinci kümede, tamamı ekstrem nokta özelliğine sahip 110 nokta (Güneybatı-Kuzeydoğu profilleri boyunca belirlenen ekstrem noktaların tamamı) olduğu görülmüştür.

Sonuç olarak, ekstrem olmayan noktaların tamamı bir kümede, ekstrem noktalar ise geriye kalan dört kümede toplanmıştır. Bu nedenle, ikinci kümedeki noktalar “Elenenler”, birinci, üçüncü, dördüncü ve beşinci kümedeki noktalar ise “Seçilenler” olmak üzere iki kümede toplanmıştır. Seçilenler kümesindeki noktalar Şekil 4.8’de görünmektedir.



Şekil 4.8 Ağırlıklandırılmış öznitelikler ile hiyerarşik kümeleme analizi sonucu seçilen, Batı-Doğu (kırmızı), Güney-Kuzey (mavi), Güneybatı-Kuzeydoğu (yeşil) ve Güneydoğu-Kuzeybatı (gök mavisi) profilleri boyunca belirlenen ekstrem grid noktaları

4.4 K-Ortalamlar Yöntemiyle Kümeleme

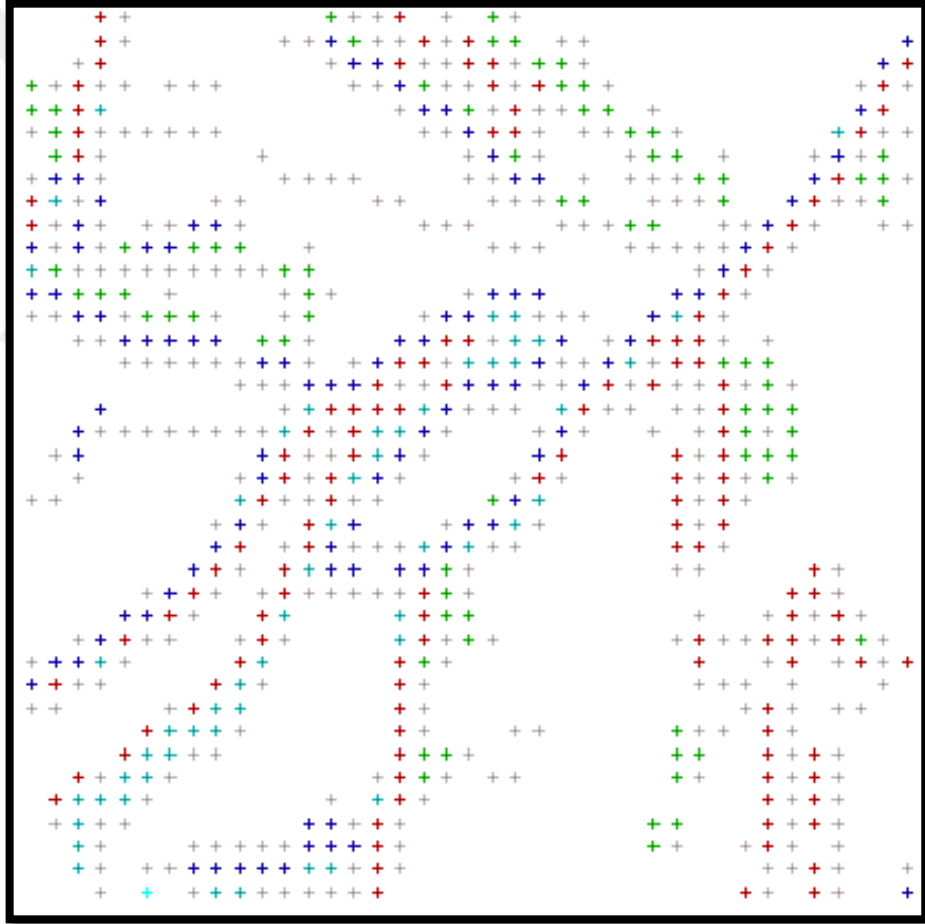
Hem eşit ağırlık öznitelikler kullanılarak hem de ağırlıklandırılmış öznitelikler ile yapılan kümeleme analizi işlemi SPSS uygulaması kullanılarak yapılmıştır. Her iki kümeleme işleminde de yakınlık ölçütü olarak Öklid mesafesi kullanılmıştır. Küme sayısı 2, maksimum iterasyon sayısı 10 ve yakınsama kriteri 0.2 olarak belirlenmiştir.

4.4.1 Eşit Ağırlıklı Öznitelikler İle Kümeleme

Kümeleme işlemi sonucunda oluşan kümeler incelendiğinde,

- Birinci kümede, 62'si ekstrem nokta özelliğine sahip toplam 803 nokta ve
- İkinci kümede, 404'ü ekstrem nokta özelliğine sahip nokta, 2'si en yüksek ve en düşük kotlu nokta olmak üzere toplam 718 nokta olduğu görülmüştür.

Sonuç olarak, ekstreme noktaların fazla olduğu ve aynı zamanda en yüksek ve en düşük kotlu noktaları da içeren ikinci küme “Seçilenler”, diğeri ise “Elenenler” kümesi olarak kabul edilmiştir. Seçilenler kümesindeki noktalar Şekil 4.9'da görünmektedir.



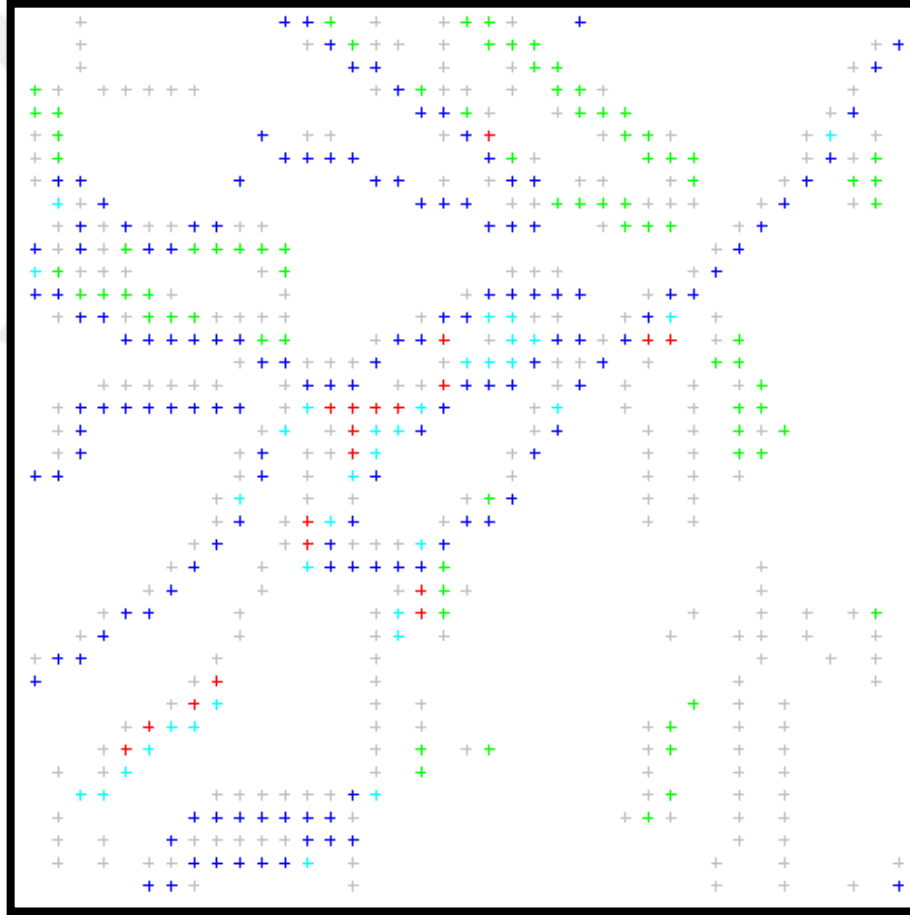
Şekil 4.9 Eşit ağırlıklı öznitelikler ile K-Ortalamalar yöntemi ile yapılan kümeleme analizi sonucu seçilen, Batı-Doğu (kırmızı), Güney-Kuzey (mavi), Güneybatı-Kuzeydoğu (yeşil) ve Güneydoğu-Kuzeybatı (gök mavisi) profilleri boyunca belirlenen ekstrem grid noktaları, ekstrem nokta olmayıp seçilen grid noktaları (gri)

4.4.2 Ağırlıklandırılmış Öznitelikler İle Kümeleme

Kümeleme işlemi sonucunda oluşan kümeler incelendiğinde,

- Birinci kümede, 187'si ekstrem nokta özelliğine sahip toplam 995 nokta
- İkinci kümede, 279'u ekstrem nokta özelliğine sahip nokta, 2'si en yüksek ve en düşük kotlu nokta olmak üzere toplam 526 nokta olduğu görülmüştür.

Sonuç olarak, ekstreme noktaların fazla olduğu ve aynı zamanda en yüksek ve en düşük kotlu noktaları da içeren ikinci küme "Seçilenler", diğeri ise "Elenenler" kümesi olarak kabul edilmiştir. Seçilenler kümesindeki noktalar Şekil 4.10'da görünmektedir.



Şekil 4.10 Ağırlıklandırılmış öznitelikler ile K-Ortalamalar yöntemi ile yapılan kümeleme analizi sonucu seçilen, Batı-Doğu (kırmızı), Güney-Kuzey (mavi), Güneybatı-Kuzeydoğu (yeşil) ve Güneydoğu-Kuzeybatı (gök mavisi) profilleri boyunca belirlenen ekstrem grid noktaları, ekstrem nokta olmayıp seçilen grid noktaları (gri)

Bu çalışmada, grid temelli SYM geneleştirilmesi amacı doğrultusunda, önemli grid noktalarının seçilmesine çalışılmıştır. Önemli grid noktalarının seçilmesi, bir yapay zekâ problemi olarak ele alınmıştır. Grid noktaları, hiyerarşik ve k-ortalama yöntemleri kullanılarak, belirlenen on üç özniteliğe göre, “Seçilenler” ve “Elenenler” olmak üzere iki kümede toplanmaya çalışılmıştır.

Öznitelikler eşit ağırlıklı Kabul edilerek ve önem derecelerine göre ağırlandırılarak her bir yönleme göre ikişer uygulama yapılmıştır. Uygulamaların başlıca sonuçları (elde edilen kümeler, küme içerikleri ve bu çalışmanın amacı doğrultusunda “Seçilenler” ve “Elenenler” olarak belirlenenler) Tablo 5.1’de verilmiştir

Tablo 5.1 Uygulamaların başlıca sonuçları

Yöntem	Öznitelik Ağırlığı	Küme	Nokta Özelliği	Nokta Sayısı	Toplam	Seçilen	Elenen
Hiyerarşik	Eşit Ağırlıklı	1	Ekstrem	414+1 (ED)	742	✓	
			Diğer	326+1 (EY)			
		2	Ekstrem	51	779		✓
			Diğer	728			
	Ağırlıklandırılmış	1	Ekstrem	144+1 (ED)	146	✓	
			Diğer	1 (EY)			
		2	Ekstrem	0	1054		✓
			Diğer	1054			
		3	Ekstrem	62	62	✓	
			Diğer	0			
		4	Ekstrem	149	149	✓	
			Diğer	0			
		5	Ekstrem	110	110	✓	
			Diğer	0			
K-ortalama	Eşit Ağırlıklı	1	Ekstrem	62	803		✓
			Diğer	741			
		2	Ekstrem	403+1 (ED)	718	✓	
			Diğer	313+1 (EY)			
	Ağırlıklandırılmış	1	Ekstrem	187	995		✓
			Diğer	808			
		2	Ekstrem	278+1 (ED)	526	✓	
			Diğer	246+1 (EY)			

Ağırlıklandırılmış özniteliklerle hiyerarşik kümeleme sonucu ortaya çıkan kümeler hariç, diğer kümelerin ekstrem nokta bakımından homojen olmadığı görülmüştür. Başka bir deyişle, ağırlıklandırılmış özniteliklerle hiyerarşik kümeleme sonucu ortaya çıkan kümeler ya bütünüyle ekstrem nokta özelliğine sahip grid noktalarından ya da bütünüyle ekstrem nokta özelliğine sahip olmayan grid

noktalarından meydana gelmişken, diğer yaklaşımlarla elde edilen kümelerde böylesi bir gruplaşma olmamıştır. Ağırlıklandırılmış özniteliklerle hiyerarşik kümeleme sonucu ortaya çıkan kümelerin ekstrem nokta bakımından homojen olması, ekstrem noktalar ile önemli noktalar arasında bir ilişki olduğunu göstermektedir.

Ağırlıklandırılmış öznitelikler kullanılarak k-ortalama yöntemine göre elde edilen kümelerin ise ekstrem nokta bakımından heterojenliği en yüksek kümeler olduğu görülmüştür.

Sonuç olarak, grid temelli SYM genelleştirilirken, önemli grid noktalarını belirlemede kullanılacak bir kümeleme yönteminden, ekstrem noktaların tamamını seçmeye elverişli kümeler oluşturması beklenir. Ancak, bu çalışmada, ele alınan iki yöntemden yalnız biri (hiyerarşik kümeleme) bu beklentiyi karşılayabilmiştir. Bununla birlikte, hiyerarşik kümeleme yönteminde de bu sonuç kendiliğinden değil, her defasında farklı bir küme sayısı göre yapılan bir dizi uygulamayla elde edilmiştir. Bu durum, hiyerarşik kümelemenin öznel (kullanıcıya bağlı) bir yapay zeka yöntemi olduğunu göstermiştir.

- Ai, T., ve Li, J., (2010), A DEM generalization by minnor valley branch detection and grid filling, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no 2, pp. 198-207.
- Akın, Y.K., (2008), Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, Doktora Tezi, İstanbul
- Alpar, R., (2013), Çok değişkenli istatistiksel yöntemler, Detay Yayıncılık, Ankara, ISBN:978-605-5437-42-8.
- Anderberg, M. R. (1973) *Cluster Analysis for Applications*, Academic Press, New York.
- Andrew, A.M., (1991), *Artificial Intelligence*, Addison-Wesley Company
- Atbaş A. C. G.,(2008), Kümeleme Analizinde Küme Sayısının Belirlenmesi Üzerine Bir Çalışma, Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Bjørke, J. T., Nilsen, S., (2003) Wavelets applied to simplification of digital terrain models, *International journal of geographical information science*, pp.601-621.
- Blashfield, R.K., Aldenderfer, M.S. (1978) The literature on cluster analysis, *Multivariate Behavioral Research*,13, 271-295
- Burmaoğlu, S., (2011). “Satınalma Alternatiflerinin Çok Değişkenli İstatistiksel Yöntemlerle Belirlenmesi: Keskin Nişancı Tüfekleri Üzerine bir Uygulama”. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*. 15: 369-382
- Charniak, E., ve McDermot, D., (1985), *Introduction to Artificial Intelligence*, Addison-Wesley Company
- Chen, Z.T., Guevara, J.A., (1987), Systematic Selection of Very Important Points (VIP) From Digital Terrain Model for Constructing Triangular Irregular Networks, in *Proceedings of the Eighth International Symposium on ComputerAssisted Cartography (AutoCarto 8)*; Baltimore, Maryland, USA.
- Chen, Z., (1989), A Spatial low-pass filter working on triangular irregular network (TIN) and restricted by break lines. In: *Proceedings of Auto-Carto 9*, Baltimore, Maryland, USA, pp. 138–145
- Copeland, J., (1993), *Artificial Intelligence: A Philosophical Introduction*, Blackwell, Oxford
- Çakmak, Z., (1999), Kümeleme analizinde geçerlilik problemi ve kümeleme sonuçlarının değerlendirilmesi, *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*. 3: 187-205.
- Çağlar, B., (2017), Mekansal verilerin kümeleme analizi ile değerlendirilmesi, Necmettin Erbakan Üniversitesi Fen bilimler Enstitüsü, yüksek lisans tezi, Konya

- Demiralay, M., Çamurcu, Y.,(2005), Cure, Agnes ve K-means Algoritmalarındaki Kümeleme Yeteneklerinin Karşılaştırılması, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 8(2): 1-18
- Dinler, M.,(2014) Kümeleme Analizi Yöntemlerinin Hayvancılık Verilerinde Karşılaştırılmalı Olarak İncelenmesi, Yüksek Lisans Tezi, Bingöl Üniversitesi, Fen Bilimleri Enstitüsü, Bingöl
- Elden, M., (2003), Hedef kitle davranışlarını etkileyen psikolojik bir faktör olarak öğrenme: Öğrenme ve reklam ilişkisi, Ege Üniversitesi, İletişim Fakültesi, İzmir
- Erden, M., Akman, Y.,(2001), Gelişim ve öğrenme, Arkadaş yayınları, Ankara
- Erilli, N.A., (2009), Kümeleme Analizine Bulanık Yaklaşım Algoritmaları Ve Uygulamaları, 19 Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi(Yayınlanmamış), Samsun.
- Erişoğlu, M., (2011), Uzaklık Ölçülerinin Kümeleme Analizine olan Etkilerinin İncelenmesi ve Geliştirilmesi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi(Yayınlanmamış), Adana.
- Everitt, S.B., Landau, S., Leese, M.,(2001) Cluster Analysis, Oxford University Press Inc, 122, New York.
- Everitt, B., (1993), Cluster analysis, (3rd edition), Halsted Press, New York, USA.
- Fırat, S. Ü., (1997). “Kümeleme Analizi: İstihdamın Sektörel Yapısı Açısından Avrupa Ülkelerinin Karşılaştırılması”. İ.Ü. Sosyal Bilimler Dergisi. III(2): 50-59.
- Fırat, F., (2016), Benzerlik ölçüleri ve kümeleme analizi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, Yüksek lisans tezi, İstanbul
- Gan, G., Ma, C., Wu, J., (2007), Data Clustering Theory Algorithms, and Applications, ASA-SIAM, Philadelphia, USA, ISBN: 978-0-898716-23-8.
- Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., (2006), Multivariate Data Analysis, Pearson Prentice Hall, USA.
- Han, J. ve Kamber, M., (2006), Data Mining: Concepts and Techniques, San Francisco.
- Hartigan, J. A., (1975). Clustering Algorithms. New York: Wiley
- Hubert, L. J., ve Levin, J. R., (1976), A General Statistical Framework for Assessing Categorical Clustering in Free Recall, Psychological Bulletin. 83(6): 1072–1080.
- Kalaycı, Ş., (Ed.), (2006), SPSS uygulamalı çok değişkenli istatistik teknikleri, Asil Yayın Dağıtım Evi, Ankara, TR, 426ss.
- Karaoğlu, Y., (2018), Farklı kümeleme analizi yöntemleri ile Türkiye’deki illerin sosyo-ekonomik özelliklerine göre sınıflandırılması ve karşılaştırılması, Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek lisans tezi, Sivas
- Kaufman L., Rousseeuw P.J.,(1990), Finding Groups in Data: An Introduction to Cluster Analysis, Wiley-Interscience, New York
- Kaya, H., Köymen, K., (2008), Veri Madenciliği Kavramı ve Uygulama Alanları, Doğu Bölgesi Araştırmaları.

- Koldere A. Y., (2008), Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, Doktora tezi, İstanbul
- Lattin, J., Carroll, D., ve Green, P., E. (2003). Analyzing Multivariate Data. Canada: Thomson Learning.
- Li, Z., (1998), An algorithm for compressing digital contour data, The Cartographic Journal, 25(2), pp.. 143-146
- Makineci, H.B., (2016), Sentinel-1A radar uydu görüntülerinden elde edilen sayısal yükseklik modelinin kalite değerlendirilmesi, Selçuk Üniversitesi Fen Bilimler Enstitüsü, Yüksek lisans tezi, Konya
- Malkoç, H., (2018), Kümeleme analizinde yapay sinir ağlarının küme geçerlilik indeksi olarak kullanılması, Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek lisans tezi, Sivas
- Mao J. ve Jain A. K., (1996). A Self-Organizing Network For Hyperellipsoidal Clustering (HEC). IEEE Transactions on Neural Networks, 7: 16-29.
- Nabiyev, V.V., (Ed.), (2016), Yapay zeka: İnsan bilgisayar etkileşimi, Seçkin Yayıncılık, Ankara, 792ss
- Özdamar, K., (Ed.), (1999), Paket programlar ile istatistiksel veri analizi, Kaan Kitabevi, Eskişehir
- Pirim, H., (2006) Yapay zeka, Endüstri Mühendisliği Bölümü, Mühendislik Mimarlık Fakültesi, Yaşar Üniversitesi
- Poorten, P., Jones, C.B., 1999. Customisable line generalization using Delaunay triangulation. In: CD-Rom, Proceedings of the 19th ICC, Ottawa, Canada,
- Pospelov D.A., (Ed.), Yapay zeka: Modeller ve yöntemler, Kitap 2, Moskova, Radio i Svyaz., 364ss
- Schmid, G., Mcgibbon, (1995), Generalization of digital terrain models for use in landform mapping , Cartographica, pp. 26-38
- Selanik, M., (2007), Türk tarımının Avrupa Birliği içindeki yerinin kümeleme analizi ile belirlenmesi, Gazi Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek lisans tezi(Yayınlanmamış), Ankara.
- Sharma S., (1996), Applied Multivariate Techniques, John Wiley and Sons, Newyork
- Silahtaroglu, G.,(Ed) (2013), Veri Madenciliği (Kavram ve Algoritmaları), Papatya Yayıncılık, İstanbul.
- Şentürk, A., (1995), Kümeleme Analizi ve Bir Uygulama, Uludağ Üniversitesi, Yüksek Lisans Tezi(Yayınlanmamış), Bursa.
- Tatlıdil, H., (Ed.), (1996) , Uygulamalı Çok Değişkenli İstatistiksel Analiz , Akademi Matbaası, 329-343, Ankara.
- Vattani, A., (2011), K-Means Requires Exponential Many Iterations Even in The Plane, Discrete Computer Geom Cilt: 45, 596-616.
- Wälder, O., Buchroithner, M. P.,(2006), Statistical modelling with locally adaptive sequential wavelet-splines, Cartographica, pp. 267-274

- Weibel, R., (1992), Concepts and experiments for adaptive computer-assisted terrain generalization, Cartography and Geographic Information System, pp. 133-153
- Wu, F., (2003), Scale – dependent representations of relief based on wavelet analysis, Geo – Spatial information science, pp. 66-69
- Yastıklı, N., Esirtgen, F., (2011), Sayısal yükseklik modellerinde kalite değerlendirme ve doğruluk analizi, TMMOB Harita ve Kadastro Mühendisleri Odası 13. Türkiye Harita Bilimsel ve Teknik Kurultayı, Ankara
- Yaz, H.F., (2014), Çok değişkenli istatistiksel tekniklerden kümeleme analizi; SPSS ile bir uygulama
- Yazgan, E., Kayaalp, G.T., (2002), Kümeleme (Cluster) Analizi Yöntemlerinin Karşılaştırmalı Olarak İncelenmesi ve Tarımsal Araştırmalarda Kullanılması, Çukurova Üniversitesi Ziraat Fakültesi Dergisi, 17 (3): 7- 14,
- URL1:<https://bilgihanem.com/ogrenme-turleri-nelerdir-ve-surecleri-nasildir/> [Ziyaret tarihi: 18 Ağustos 2019]
- URL2:<http://www.psikolojibilimi.gen.tr/ogrenme-nedir-psikolojide-ogrenme-turleri-nelerdir/> [Ziyaret tarihi: 18 Ağustos 2019]

TEZDEN ÜRETİLMİŞ YAYINLAR

İletişim Bilgisi: a.kirmizibiber@atauni.edu.tr

Konferans Bildirileri

1. Kırmızıbiber, A., Gökgöz, T.,(2019), Hiyerarşik ve k-ortalamalar yöntemleriyle grid noktalarının kümelenmesi, TMMOB Harita ve Kadastro Mühendisleri Odası, 17. Türkiye Harita Bilimsel ve Teknik Kurultayı, Ankara

