Bulanık Kümeleme Analizi ve OECD Ülkelerinin Gelişmişlik Bakımından Kümelendirilmesi

Serra Atal

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

Ağustos 2015

Fuzzy Cluster Analysis and Clustering OECD Countries Based on Development

Serra Atal

MASTER OF SCIENCE THESIS

Department of Statistics

August 2015

Bulanık Kümeleme Analizi ve OECD Ülkelerinin Gelişmişlik Bakımından Kümelendirilmesi

Serra Atal

Eskişehir Osmangazi Üniversitesi
Fen Bilimleri Enstitüsü
Lisansüstü Yönetmeliği Uyarınca
İstatistik Anabilim Dalı
Uygulamalı İstatistik Bilim Dalında
YÜKSEK LİSANS TEZİ
Olarak Hazırlanmıştır

Danışman: Prof. Dr. Zeki Yıldız

ONAY

İstatistik Anabilim Dalı Yüksek Lisans öğrencisi Serra Atal'ın YÜKSEK LİSANS tezi olarak hazırladığı "Bulanık Kümeleme Analizi ve Bir Uygulama" başlıklı bu çalışma, jürimizce lisansüstü yönetmeliğin ilgili maddeleri uyarınca değerlendirilerek oybirliği ile kabul edilmiştir.

Danışman **:** Prof. Dr. Zeki Yıldız

İkinci Danışman **:** -

Yüksek Lisans Tez Savunma Jürisi:

Üye : Prof. Dr. Zeki YILDIZ

Üye : Prof. Dr. Veysel YILMAZ

Üye : Doç. Dr. Sevil ŞENTÜRK

Üye : Y. Doç. Dr. Özer ÖZAYDIN

Üye : Y. Doç. Dr. Fatih ÇEMREK

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

Prof. Dr. Hürriyet ERŞAHAN
Enstitü Müdürü

ETİK BEYAN

Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna göre, Prof. Dr. Zeki Yıldız danışmanlığında hazırlamış olduğum "Bulanık Kümeleme Analizi ve OECD Ülkelerinin Gelişmişlik Bakımından Kümelendirilmesi" başlıklı YÜKSEK LİSANS tezimin özgün bir çalışma olduğunu; tez çalışmamın tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; tezimde verdiğim bilgileri, verileri akademik ve bilimsel etik ilke ve kurallara uygun olarak elde ettiğimi; tez çalışmamda yararlandığım eserlerin tümüne atıf yaptığımı ve kaynak gösterdiğimi ve bilgi, belge ve sonuçları bilimsel etik ilke ve kurallara göre sunduğumu beyan ederim. 10/08/2015

Serra Atal

İmza

İÇİNDEKİLER

Sayfa

İÇİNDEKİLER (devam ediyor)

İÇİNDEKİLER (devam ediyor)

ŞEKİLLER DİZİNİ

Şekil Sayfa

ÇİZELGELER DİZİNİ

SİMGELER VE KISALTMALAR DİZİNİ

Simgeler	Açıklama
$\|\| \quad \|\|$	Vektör Normu
$\forall$	Her
$\in$	Eleman
D	Uzaklıklar Matrisi
det	Determinant
J	Amaç Fonksiyonu
max	Maksimum
min	Minimum
S	Kovaryans Matrisi
V	Küme Prototipi
μ	Mü
ρ	Rho

Kısaltmalar	Açıklama
PC	Bölünme Katsayısı
CE	Sınıflandırma Entropisi
OECD	Ekononomik İşbirliği ve Kalkınma Örgütü

ÖZET

Kümeleme yöntemleri; verinin kümelere nasıl atandıklarına, başka bir deyişle, hangi türde bölünmeler oluşturduklarına göre ayrılırlar. Klasik kümeleme yöntemlerinde her birim kesin olarak bir kümeye atanmak zorundadır. Bu yöntemler, veri setini eksiksiz olarak boş olmayan ve ikili ayrık alt gruba ayrıştıran bölünmeler üretir. Zadeh tarafından geliştirilen bulanık küme teorisi ile açıklanan "üyelik fonksiyonu" ile kesin olarak ait olma durumu ortadan kalkmış ve klasik kümelemeye alternatif olan Bulanık Kümeleme Yöntemi ortaya çıkmıştır. Bulanık kümeleme yönteminin kullanımı, belirsiz küme üyelikleri hakkında bilgi sağlar. Bulanık kümeleme, her bir birimin sadece tek bir kümeye atanma zorunluğunu ortadan kaldırarak, her bir birimin belli üyelik dereceleriyle tüm kümelere üye olduğu bir kümeleme yöntemine dönüştürür.

Bu tez çalışmasında Ekononomik İşbirliği ve Kalkınma Örgütü (OECD- The Organization for Economic Co-operation and Development) üyesi 34 ülkenin sosyo-ekonomik yapılarını gösteren 30 değişkene ait veriler kullanılmıştır. Ülkelerin ekonomik performansını gösteren makro-ekonomik ve sosyo-kültürel göstergeler ile örgütlenmede yer alan bu ülkelerin gelişmişlik bakımından hangileriyle benzeştiğinin, hangileriyle farklılaşma gösterdiğinin ortaya konulması hedeflenmiştir. Bu amaçla bulanık kümeleme analizinden yararlanılmıştır. Bulanık kümeleme analizi hem MATLAB hem de R paket programları kullanılarak gerçekleştirilmiş, sonuçlar karşılaştırılmıştır. Ek olarak aynı veriye klasik kümeleme yöntemlerinden olan k-medoid kümeleme yöntemi uygulanarak bulanık kümeleme sonuçları ile karşılaştırılmıştır. Uygulama sonucunda bulanık kümeleme ve k-medoid kümeleme yöntemleri ile OECD ülkeleri gelişmiş ve daha az gelişmiş olmak üzere iki ayrı kümeye ayrılmıştır. Ülkelerin yer aldıkları kümeler bulanık kümeleme yöntemi ve k-medoid kümeleme yöntemine göre farklılık göstermiştir.

Anahtar Kelimeler: kümeleme, bulanık mantık, bulanık kümeleme, bulanık c-ortalamalar

SUMMARY

Clustering methods split up according to how data is assigned to clusters, in other words, what kinds of groups they create. In classical clustering methods, each unit certainly has to be assigned to one cluster. These methods generate the groups which seperate the data set to completely non-empty and dual discrete subgroup. With “Membership Function” described by the fuzzy set theory developed by Zadeh, definitely belonging situation disappeared and Fuzzy Clustering Method which is an alternative to classical clustering arose. The use of fuzzy clustering method provides information about improper cluster memberships. Fuzzy clustering eliminates the necessity that each unit is assigned to only one cluster, converts to a clustering method which each unit is a member of all clusters with certain degrees of membership.

In this thesis study, the datas belonging to 30 variables which shows the socio-economic structures of 34 countries that are members of The Organization for Economic Co-operation and Development(OECD) were used. Revealing that these countries in the organization are similar to which of these, shows differentiation with which of these in terms of development with macro-economic and socio-cultural indicators showing economic performances of countries was aimed. For this purpose, fuzzy clustering analysis was used. Fuzzy clustering analysis was performed by using both MATLAB and R packaged softwares, the results were compared. Additionally, k-medoids clustering method which is one of classical clustering methods was applied to the same data, compared with the results of fuzzy clustering. As a result of the application, OECD countries were divided into two seperate clusters including developed and least developed with fuzzy clustering and K-medoids clustering. The clusters which the countries located in varied depending on fuzzy clustering method and k-medoids clustering method.

Keywords: clustering, fuzzy logic, fuzzy clustering, fuzzy c-means

TEŞEKKÜR

Tez çalışmam boyunca bilgi ve tecrübelerini benimle paylaşarak bana yol gösteren sayın danışmanım Prof. Dr. Zeki Yıldız’ a,

Yüksek lisans eğitimim süresince bilgilerinden yararlandığım ve bana her konuda yardımcı olan değerli istatistik bölümü hocalarıma ve benden yardımını esirgemeyen sayın Doç. Dr. Sevil Şentürk’ e,

Her zaman yanımda olduklarını bildiğim aileme ve arkadaşlarıma,

sonsuz teşekkürlerimi ve şükranlarımı sunarım.

1. GİRİŞ

Nesnelerin ve olayların sınıflandırılması insanların temel faaliyetlerinden biridir. Biz insanlar, belki de karmaşık dünyayı basitleştirme çabasıyla yaşadığımız çevreyi anlamlı birimlere bölme eğilimindeyizdir. Örneğin; cinsiyet, siyasi partiler veya nesne ya da olayları özellik ve kullanımlarına göre sınıflara-türlere ayırmak gibi (Tinsley ve Brown, 2000).

Bilimsel araştırmaların çoğunun en önemli unsuru üzerinde çalışılan olayın sınıflandırılmasıdır. Örneğin, davranışsal bilimlerde, sınıflandırılan, bireyler ya da toplumlar ya da davranış ve algı kalıpları dahi olabilir. Araştırmacı genellikle ilgili öğelerin az sayıda homojen grup ya da kümede yer aldığı sınıflandırmayı bulmakla ilgilenir. Bu grup ya da kümeler genellikle birbiriyle bağdaşmaz ama bu bir zorunluluk değildir. Sınıflama en azından, üzerinde sınıflama yapılan çok değişkenli verinin işe yarar bir özetini sağlayabilmelidir, ama çoğunlukla bundan daha fazlası elde edilir (Everitt ve Dunn, 1991).

İstatiksel sınıflandırma yöntemlerinden biri olan kümeleme analizi, karmaşık yapıdaki olayların ya da nesnelerin sınıflandırılmasını ayrıntılı biçimde açıklayan bir yöntemdir. Klasik bir sınıflama yöntemi olan kümeleme analizi, son yıllarda geliştirilen metotlarla birleşerek daha farklı ve daha geniş bir bakış açısına sahip olmuştur. Bulanık kümeleme analizi ise; kümeleme analizinin bulanık mantık ile birlikte genişletilmesiyle oluşan ve uygulama alanı oldukça yaygın olan alternatif bir yöntemdir.

Çalışmada, bulanık kümeleme analizi ile Türkiye'nin üyesi bulunduğu OECD (The Organization for Economic Co-operation and Development- Ekonomik İşbirliği ve Kalkınma Örgütü) ülkeleri arasında sosyoekonomik gelişmişlik bakımından nerede olduğu, hangi ülkeler ile benzeştiği araştırılmıştır. Bilindiği üzere, OECD, 1961 yılında kurulmuş olup çoğu Avrupa Birliği ülkelerinden oluşmaktadır. Sonradan katılan üyelerle birlikte, şu an 34 üyeye sahiptir. Bu araştırma, Avrupa Birliği sürecinde olan Türkiye'nin OECD'de yer alan mevcut AB ülkeleri ve aday olan ülkeler ile ne kadar ayrışıp benzeştiğini ortaya koymayı hedeflemektedir.

Bu amaçla çalışmanın ikinci bölümünde ilgili literatür araştırmasına, üçüncü bölümünde kümeleme analizine dair teorik bilgiye, dördüncü bölümünde bulanık kümelemeye dair teorik bilgiye, beşinci bölümünde ise uygulamaya yer verilmiştir. Altıncı bölümde uygulamaya dair bulgular ve tartışmaya ve son olarak da yedinci bölümde sonuç ve önerilere yer verilerek çalışma tamamlanmıştır.

2. LİTERATÜR ARAŞTIRMASI

Bu çalışmada ülkelere ait sosyoekonomik verilerden yararlanılarak kümeleme analizi gerçekleştirilmiştir. Literatürde, gerek toplumdaki bireylere, gerek illere, gerekse ülkelere vb. topluluklara ait demografik ve sosyoekonomik veriler kullanılarak kümeleme analizi ile bireylerin, illerin, ülkelerin, vb. gruplanmalarını ortaya çıkarmak üzere yapılan çalışmalar mevcuttur.

Yıldız (1989), banka müşterilerinin demografik ve sosyoekonomik özelliklerine ait veriler yardımıyla kümeleme analizi uygulamış ve müşteri kalıplarının belirlenmesine yardımcı olacak çeşitli gruplara ayırmıştır. Aşan (2007), ise benzer biçimde kredi kartı kullanıcılarına ait sosyoekonomik verilerle kümeleme analizi uygulamıştır. Gökgöz vd. (2013) ise Türk bankacılık sistemindeki bankalara ait finansal değişkenleri kullanarak bulanık kümeleme analiziyle bankaları gruplandırmıştır.

Sönmez ve Er (2006), Türkiye'deki illere ait demografik verileri kullanarak modern kümeleme yöntemleri ile iç göç hareketleri bakımından benzer illeri tespit etmiştir. Erilli vd. (2008), illeri sosyoekonomik verilere dayanarak bulanık kümeleme analizi ile sınıflandırmıştır. Atalay ve Tortum (2010), Türkiye'deki illere ait trafik kaza verileri ile geleneksel ve bulanık küme analizlerini kullanan karşılaştırmalı analizlerle illeri kümelere ayırmıştır. Yılancı (2010), Türkiye'deki illere ait sosyoekonomik verilere bulanık kümeleme analizini uygulamıştır. Kılıç vd. (2012) ise Türkiye'deki illeri hayvancılık istatistikleri bakımından bulanık kümeleme analizi ile sınıflandırmıştır.

Terlemez (2001), çalışmasında kümeleme analizi ile AB'ye üye ülkelerin ekonomik durumunu incelemiştir. Şahin ve Hamarat (2002), G-10, Avrupa Birliği ve OECD ülkelerinin sosyoekonomik benzerliklerini ortaya koyan çalışmalarında bulanık kümeleme analizini kullanmışlardır. Çemrek vd. (2010), bulanık kümeleme analizi ile OECD ülkelerini CO_2 emisyonları bakımından incelemiştir. Yeloğlu (2009), Türkiye ile OECD ülkelerini ekonomik değişkenler kullanarak karşılaştıran çalışmasında, hiyerarşik kümeleme yöntemini uygulamıştır. Giray (2013), turizm istatistikleri bakımından 159 ülke arasında Türkiye'nin yerini belirlemek amacıyla yaptığı çalışmada hiyerarşik ve bulanık kümeleme analizlerini kullanmıştır. Akın ve Eren (2012), OECD ülkelerine ait eğitim göstergelerini kullanarak kümeleme analizi ile ülkelerin gruplanmalarını belirlemişlerdir. Giray ve Gülel (2014), bulanık kümeleme analizi ile Avrupa ülkelerini intihar oranlarına göre sınıflandırmıştır.

3. KÜMELEME ANALİZİ

Kümeleme analizi (segmentasyon analizi ya da taksanomi analizi) incelenen birimleri aralarındaki benzerliklere dayanarak birbirine göre homojen alt birimlere ayırmayı gerçekleştiren teknikler dizisidir (Kashigan, 1991).

Kümeleme analizi, çok değişkenli verideki homojen ve diğerlerinden ayrı grupları/ kümeleri ortaya çıkarmak ya da anlamak için inceleyen çeşitli sayısal yöntemlerin genel adıdır (Everitt ve Hothorn, 2006).

Kümeleme analizi, Linnaeus (1753) ‘un hayvanların ve bitkilerin sınıflandırılması üzerine yaptığı çalışmaya dayanan bir araştırma yöntemidir (Hofman ve Jarvis, 1998).

Biyolojide organizmaların sınıflandırılmasına ait teori ve uygulamalar *taksanomi* adıyla bilinir. Başlangıçta taksanominin etkisi bilimsel metottan daha çok sanat alanında görülse de çoğunluğu Adanson (1727-1806) tarafından geliştirilen tekniklerle daha geniş kullanım alanına sahip olmuştur. Sneath ve Sokal (1963) bu tekniklere dayandırdığı *“Numerical Taxonomy”* adlı çalışmada birimleri çeşitli özelliklerine göre sınıflandırmışlardır (Everitt, 1993).

Sınıflandırma oluşturmak için kullanılan sayısal teknikler, biyoloji ve zooloji gibi fen bilimlerinde uygulanışıyla ortaya çıkmıştır. 20.yy.’ın yarısından itibaren ise bu tekniklerin kullanımı önemli ölçüde artmıştır. Bu sayısal teknikler, uygulandıkları alanlara göre çeşitli isimlere anılmaktadır. *“Sayısal Taksanomi”* terimi genellikle biyolojide kullanılırken, psikolojide “Q Analizi”, yapay zeka literatüründe “Güdümsüz Örüntü Tanıma” terimleri kullanılmaktadır. Bununla birlikte en genel terim ise “Kümeleme Analizi” dir (Everitt, 1993).

Kümeleme analizinin uygulandığı çeşitli alanlar vardır. Örnek olarak; bireyleri sosyal tutum, özgüven, kan değerleri, hastalık geçmişi ya da müşteri gereksinimi benzerliklerine dayanarak kümeleyebiliriz. Bitki ve hayvan numuneleri ya da tüm önceden tanımlanmış türler, çeşitli morfolojik, psikolojik ya da çevresel karakterlerine dayanarak kümelenebilir. Benzer olarak, ırklar, dinler, kültürler, maden örnekleri, fosiller, arkeolojik eserler birbirlerine benzerliklerine göre gruplara kümelenebilir. Kısacası, herhangi bir nesne bütünü, kümeleme analizinin konusu olabilir (Kashigan, 1991).

Son yıllarda ise kümeleme analizi teknikleri, microarray verilere (Alon vd. , 1999), görüntü analizi (Everitt ve Bullmore, 1999) ve pazarlama biliminde (Dolnicar ve Leisch, 2003) uygulanmaktadır (Everitt ve Hothorn, 2006).

Kümeleme analizi tekniği, gözlemleri gruplara ya da kümelere toplamakta kullanılır;

i. Her grup ya da küme belirli bir özelliğe göre homojendir/ düzenlidir. Her gruptaki gözlemler birbirlerine benzerdir.
ii. Her grup aynı özelliğe göre diğer gruplardan farklı olmalıdır; bir gruptaki gözlemler, diğer gruplardaki gözlemlerden farklı olmalıdır (Sharma, 1996).

Kümeleme analizi; X veri matrisinde yer alan doğal gruplamaları kesin olarak bilinmeyen bireyleri, değişkenleri ya da birey ve değişkenleri birbirleri ile benzer olan alt kümelere (grup, sınıf) ayırmaya yardımcı olan yöntemler topluluğudur (Özdamar, 2010).

Kümeleme analizinin genel amaçlarının dışında özel amaçlarını da şöyle sıralayabiliriz:

i. Gerçek tiplerin belirlenmesi
ii. Model uydurmanın kolaylaşması
iii. Gruplar için ön tahmin
iv. Hipotezlerin testi
v. Veri yapısının netleştirilmesi
vi. Veri indirgenmesi
vii. Aykırı değerlerin bulunması (Tatlıdil, 2002)

3.1. Uzaklık (Benzerlik) Ölçüleri

Kümeleme analizinin temel amacı ise birimlerin (ya da değişkenlerin) doğal gruplanmalarını ortaya çıkarabilmektir. Bunun için önce nesneler arasındaki benzerliği ölçen niceliksel bir ölçek oluşturulmalıdır (Johnson ve Wichern, 1998).

Kümeleme analizinde en önemli adım üzerinde çalışılan nesneler arasındaki benzerlik (yakınlık) ölçüsünü elde etmektir. Alternatif olarak; benzerlik ve uzaklığın birbirinin tamamlayıcısı olduğundan, nesneler arası uzaklık ya da fark ile ilgileniriz (Kashigan, 1991).

Kümeleme analizinde, birimlerin p değişkene göre birbirleri arasındaki uzaklıklarını hesaplamak için çok çeşitli uzaklık ölçü birimleri öne sürülmüştür.

Uzaklık ölçüleri ya da benzerlik ölçüleri, veri matrisinde yer alan değişkenlerin ölçü birimlerine göre de farklılık göstermektedir. Eğer değişkenler oransal ya da aralıklı ölçeklerle elde edilmiş değerler ise uzaklık (distance) ya da ilişki (correlation) türü ölçülerden yararlanılır. Ölçümler sayısal değerler olarak yapılmış ise tercih edilen ölçüler kikare uzaklık ölçüsü (chi-square measure) ya da Phi kare uzaklık ölçüsüdür (phi-square measure). Eğer ikili (binary) gözlemler göre ölçümler yapılmış ise birimler arasındaki benzerlikleri belirlemede Öklid, kare Öklid, size difference, pattern difference, Lance and Williams difference, shape difference gibi benzerlik ya da farklılık ölçülerinden yararlanılmaktadır.

Genelde uzaklık ölçüleri doğrudan birim ya da değişkenlerin kümelenmesinde kullanılacağı gibi birim ya da değişkenler arasındaki benzerlik ya da farklılıkları (similarity or dissimilarity) hesaplamakta da kullanılırlar (Özdamar, 2010).

Kümeleme analizinde sıklıkla kullanılan bazı uzaklık ölçüleri izleyen biçimdedir:

- Öklid uzaklığı: Kümeleme analizinde en kolay anlaşılabilen ve en yaygın olarak kullanılan uzaklık ölçüsü Öklid uzaklığıdır.

$$d_{ij} = \left[\sum_{k=1}^{p}(x_{ik} - x_{jk})^2\right]^{1/2}, i = 1,2,\dots,n\,;\ j = 1,2,\dots,n\ ve\ k = 1,2,\dots,p \qquad (3.1)$$

Öklid uzaklığı, noktalar ve kümeler arasındaki uzaklığı ölçmek için tek seçenek değildir ve bazı durumlarda kötü bir seçim olabilir. Özellikle veri vektörünün birimleri farklı ölçeklerle ölçülmüş birbirine benzemez birimlerse, en büyük değerlere sahip olan değişken Öklid uzaklığını domine etme eğilimindedir (Wilks, 2011).

Öklid uzaklığı, tek bir değişkenin aykırı değerlerine ekstra ağırlık verme özelliğinden dolayı bazı kullanıcılar tarafından sevilmemektedir. Standartlaştırma ile kısmen bunun üstesinden gelinir (Cormack, 1971).

Ayrıca değişkenlere göre toplam uzaklığın karekökü alınmaksızın *"Karesel Öklid Uzaklığı"* da hesaplanmaktadır.

- Minkowski uzaklığı: Bazı kümeleme yöntemlerinde uzaklığın karekökünü almak gerekmez. Bu durumlarda önerilen uzaklık ölçülerinden biri Minkowski uzaklığıdır (Alvin ve Christensen, 2012).

$$d_{ij} = \left[\sum_{k=1}^{p}|x_{ik} - x_{jk}|^{\frac{1}{r}}\right]^{r}, i = 1,2,\dots,n;\ j = 1,2,\dots,n; k = 1,2,\dots,p\ ve\ r \geq 1 \qquad (3.2)$$

Formülde r=2 olduğunda Öklid uzaklığına dönüşür.

Minkowski uzaklığı formülünde p=2 ve r=1 olduğunda *"City-Block (Manhattan) Uzaklığı"* elde edilir.

- Mahalanobis uzaklığı: Mahalanobis uzaklığı formüldeki gibidir.

$$d_{ij} = (x_i - x_k)' S^{-1} (x_i - x_k) \tag{3.3}$$

Burada S kovaryans matrisini göstermektedir.

Mahalanobis uzaklığının avantajı, sadece sürekli değişkenler arasındaki korelasyonu değil her değişkenin birimleriyle varyansını da göz önünde bulundurmasıdır. Burada araştırmacının birimleri bireylerden oluşuyorsa muhtemelen buna ihtiyacı olmayacaktır. Ama ülkeler, şehirler, kurumlar vb. sosyal topluluklarla çalışan araştırmacılar kesinlikle Mahalanobis uzaklığını dikkate almak isteyeceklerdir (Bailey, 1975).

- Canberra ölçüsü: Sadece negatif olmayan değişkenler için tanımlanmıştır (Johnson ve Wichern, 1998) :

$$d_{ij} = d(x_i, y_j) = \sum_{i=1}^{n} \frac{|x_{ki} - y_{kj}|}{|x_{ki} + y_{kj}|}, i \neq j = 1,2, \dots, n \tag{3.4}$$

- Pearson korelasyon ölçüsü: Bilinen en eski uzaklık ölçülerinden biri olan ve birimler arası varyansı içeren Pearson korelasyon katsayısı (ırksal benzerlik ölçüsü- CRL), Pearson (1926) tarafından kafataslarını karşılaştırmak için geliştirilmiştir (Bailey,1975). Sürekli iki değişken arasındaki doğrusal ilişkiyi gösterir.

$$d_{ij} = \sqrt{\sum_{k=1}^{p} \left(x_{ik} - x_{jk}\right)^2 / S_k^2} \quad , i, j = 1,2, \dots, n; k = 1,2, \dots, p \tag{3.5}$$

3.2. Verilerin Standardizasyonu ve Dönüştürülmesi

Veri matrisinde değişkenlerin ortalama ve varyansları birbirinden çok farklı olduklarında, büyük ortalama ve varyansa sahip değişkenler diğer değişkenleri istatistiksel analizlerde belirli oranda baskılamakta ve rollerini, etkinliklerini göreceli olarak azaltmaktadır. Bazen değişkenlerin aşırı uçlarda yer alan değerleri kümeleme üzerinde olumsuz etkilerde bulunmaktadır. Bu gibi durumlarda verilerin standardize ya da belirli aralıklarda gözlenen değerlere dönüştürülmesi (transforme edilmesi) uygun olmaktadır.

Verilerin standardize edilmesi ya da belirli aralıklara dönüştürülmesi için birçok yöntem bulunmaktadır. Bu yöntemler; z skorlarına dönüştürme, $-1 \leq x \leq 1$ aralığına dönüştürme, $0 \leq x \leq 1$ aralığına dönüştürme, maksimum değer 1 olacak şekilde dönüştürme ve standart sapma 1 olacak şekilde dönüştürme gibi yöntemlerdir. (Özdamar, 2010).

3.3. Kümeleme Yöntemleri

Kümeleme yöntemleri; uzaklık matrisi ya da benzerlik matrisinden yararlanarak birimler ya da değişkenleri kendi içinde homojen ve kendi aralarında heterojen uygun gruplara ayırır (Özdamar, 2010).

Kümeleme yöntemleri;

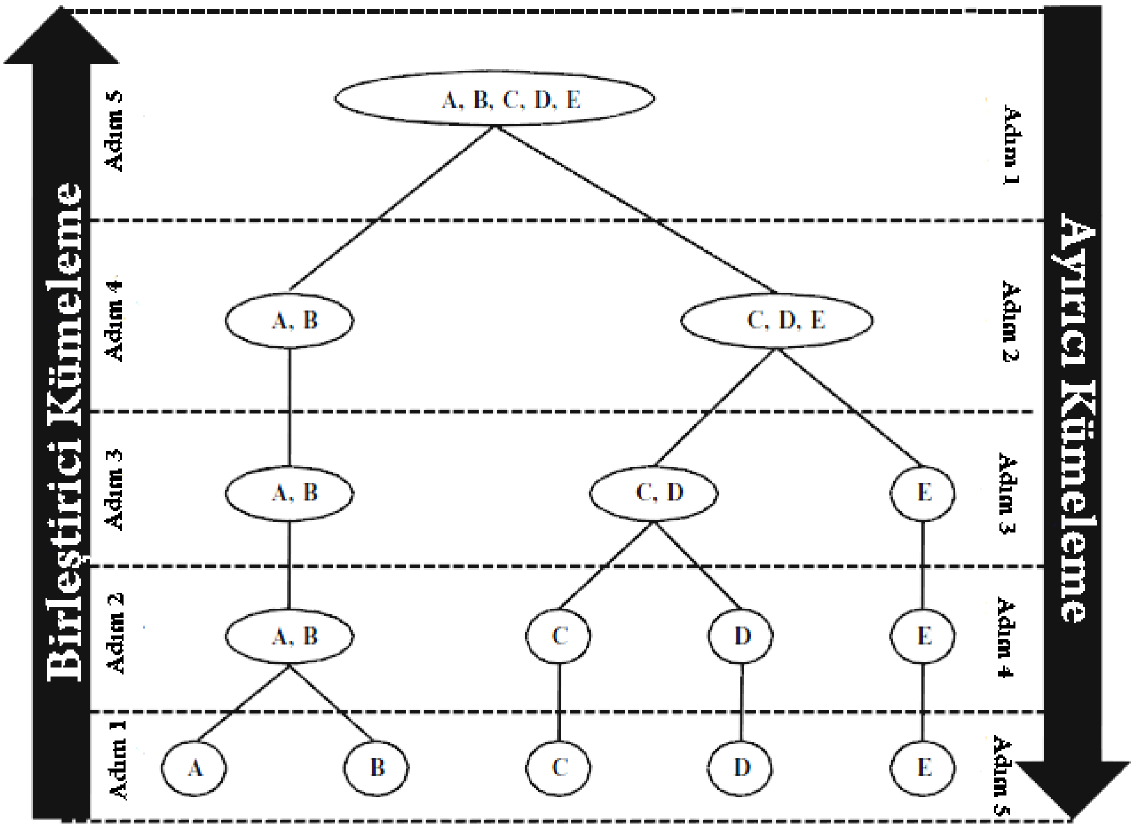
i. Hiyerarşik Kümeleme Yöntemleri
ii. Hiyerarşik Olmayan Kümeleme Yöntemleri

olmak üzere iki ana kategoriye ayrılır (Sharma, 1996).

3.3.1. Hiyerarşik kümeleme yöntemleri

Hiyerarşik yöntemler, genellikle tek birim içeren kümelerden başlayarak, tüm birimler bir kümede toplanana kadar birleştirme işlemi yaparak küme dizileri üretir. Bu yöntemler *"Birleştirici Hiyerarşik Kümeleme Yöntemi"* olarak adlandırılır. Diğer yöntemler ise; tek bir kümeyle başlayarak, birimleri art arda ayırarak, tek birim içeren kümeler oluşturana dek devam eder. Bu yöntemlere ise *"Ayırıcı Hiyerarşik Kümeleme Yöntemleri"* denir (Nell, 2002).

Hiyerarşik sınıflamalar "*dendrogram*" olarak bilinen iki boyutlu şemalarla ifade edilebilir. Dendrogramlar birbirini izleyen her bir aşamadaki birleşmeleri ya da bölünmeleri gösterir (Everitt, 1993). Birleştirici ve ayırıcı kümelemeye ilişkin örnek dendrogram Şekil 3.1' de verilmiştir.



Şekil 3. 1. Birleştirici ve Ayırıcı Kümeleme Örnek Dendrogram (Mooi ve Sarstedt, 2011)

3.3.1.1. Birleştirici hiyerarşik kümeleme yöntemleri

Birleştirici Hiyerarşik Kümeleme Yöntemi, Sokal ve Sneath (1963) tarafından "aşağıdan başlayan sınıflama", Lyerly (1967) tarafından "sentetik" olarak adlandırılırken; Gower (1967) bu yöntem için "birleştirici" kelimesini kullanmıştır. Ayırıcı Hiyerarşik Kümeleme Yöntemi için de Sokal ve Sneath (1963) "yukarıdan başlayan sınıflama" ve Lyerly (1967) "analitik" kelimelerini kullanırken; Gower (1967) bu yöntemi "ayırıcı" olarak adlandırmıştır (Bailey,1975).

n birime sahip bir örnek için birleştirici hiyerarşik kümeleme algoritması şu şekildedir:

i. Her biri tek birime sahip n kümeleri ile başlanır ve $D = \{d_{ij}\}$ uzaklıklarına (benzerliklerine) sahip simetrik $n \times n$ boyutlu matris hesaplanır.
ii. Uzaklık matrisinde en yakın (en benzer) küme çiftleri araştırılır. En benzer kümeler; U ve V arasındaki d_{UV} uzaklığı belirlenir.
iii. U ve V kümeleri birleştirilir. Yeni oluşan bu küme UV olarak adlandırılır. Benzerlik matrisinde U ve V kümelerini temsil eden satır ve sütun iptal edilir ve matristen çıkartılır. (UV) kümesi ve kalan kümeler arasındaki uzaklığı veren satır ve sütun eklenerek benzerlik matrisi güncelleştirilir.

İkinci ve üçüncü adımlar $n-1$ kez tekrarlanır. (Algoritma sona erdiğinde tüm birimler tek bir kümede olacaktır.) Birleştirilen kümelerin özellikleri ve birleşmenin yapıldığı düzeyler (uzaklıklar ya da benzerlikler) kaydedilir (Johnson ve Wichern, 1998).

Birleştirici hiyerarşik kümeleme yöntemleri birimler arasındaki uzaklıklar matrisi ile başlar. Tüm birimler tek başına birer grup kabul edilerek başlanır ve bu gruplar kendilerine "yakın" olan kümelerle birleştirilir. Bu "yakın"lığı tanımlamak için farklı yollar mevcuttur (Manly, 1994). En çok kullanılan birleştirici hiyerarşik kümeleme yöntemleri şunlardır:

i. Tek Bağlantı Kümeleme (Single Linkage Method, Nearest Neighbours Method)
ii. Ortalama Bağlantı Kümeleme (Average Linkage Method)
iii. Tam Bağlantı Kümeleme (Complete Linkage Method, Furthest Neighbours Method)
iv. Küresel Ortalama Bağlantı Kümeleme (Centroid Method)
v. Ward Hiyerarşik Kümeleme (Ward's Hierarchical Clustering)

➢ Tek Bağlantı Kümeleme (Single Linkage Method, Nearest Neighbours Method):

Bu teknik ilk olarak Florek et al. (1951) tarafından *"dendritik metod"* olarak tanımlanmıştır. McQuitty (1957) ve Sneath (1957) ve Johnson (1967) bağımsız olarak bu metodun daha farklı bir versiyonunu tanıtmışlardır (Everitt ve Dunn, 1991).

Tek bağlantı kümelemede iki küme arasındaki uzaklık olarak; iki kümenin herhangi iki elemanına en kısa uzaklık alınır (Mooi ve Sarstedt, 2011). Bu yöntemde önce birbirine en yakın iki birim (gözlem) bir kümeye yerleştirilir. Daha sonra diğer en yakın uzaklık tespit edilerek ilk oluşturulan kümeye bu gözlem eklenir veya iki gözlemden oluşan yeni bir küme oluşturulur. İşlem tüm gözlemlerin bir kümeye yerleştirilmesine kadar devam eder.

Bu yöntemde eğer i. ve j. birimler birleştirilmiş ise birleştirilen kümenin k. küme ile ilişkisi uzaklık ölçütü olarak,

$$d_{k(i,j)} = \min(d_{ki}, d_{kj}) \tag{3.6}$$

şeklinde ifade edilir.

Burada;

$d_{k(i,j)}$: k'ıncı kümenin daha önce oluşan i. ve j. kümelerle olan uzaklığını,

d_{ki}: k'ıncı kümenin j'inci kümeye olan uzaklığını,

d_{kj}: k'ıncı kümenin i'nci kümeye olan uzaklığını gösterir (Çelik,2013).

- Ortalama Bağlantı Kümeleme (Average Linkage Method):

Ortalama bağlantı kümeleme iki küme arasındaki uzaklığı; kümelerdeki nokta çiftleri arasındaki uzaklığın ortalaması olarak tanımlar. Ortalama bağlantı kümeleme, popülasyon dağılımının duyarlılığında tam bağlantı kümelemeye göre daha iyidir. Çünkü uzaklık ölçüsü kümedeki noktaların sayısından etkilenir (Hartigan,1985).

Ortalama bağlantı kümelemede bir birimin m. küme olarak hangi birim ya da kümelerle birleştirileceği, birimlerin yeni oluşan kümelerle olan uzaklıkları dikkate alınarak belirlenir. m. kümenin daha önce oluşan k. ve l. Kümelerden hangisi ile birleşeceğini belirlemek için j. küme ile k. ve l. kümelerin uzaklıklarına bakılır. Bu uzaklıklar k ve l kümelerinin eleman sayısı ile çarpılarak ağırlıklandırılır. Elde edilen toplam yeni oluşacak m. küme eleman sayısına bölünür.

m. kümenin j. küme ile olan uzaklığı (d_{mj}),

$$d_{mj} = \frac{N_k d_{kj} + N_l N_{lj}}{N_m} \tag{3.7}$$

En çok benzer çiftin bulunmasında grup içi benzerlik ortalaması k ve l kümelerine ait çiftlerin benzerlik ölçülerinden ve birim sayılarından yararlanılarak yapılır (Özdamar, 2010).

➢ Tam Bağlantı Kümeleme (Complete Linkage Method, Furthest Neighbours Method):

Tam bağlantı kümelemede; tek bağlantı kümelemenin aksine gruplar arası uzaklık her gruptaki birimlerin birbirine olan uzaklıkları arasından en uzak olan birim çifti olarak tanımlanır (Everitt,1993). Bu yöntemde uzaklık;

$$d_{k(i,j)} = \max(d_{ki}, d_{kj}) \tag{3.8}$$

şeklinde gösterilir (Çelik, 2013).

Kullanılmakta olan paket programların çoğu tek ve tam bağlantı tekniklerini kullanmaktadır. Tek bağlantı yöntemi, sağlıklı sonuçlar vermesi dolayısıyla tercih edilmekte ancak işlemlerin uzun sürmesi açısından da sakıncalıdır. Tam bağlantı tekniği ise aynı küme içerisindeki bireylerin uzaklıklarının belli bir değerden küçük olması durumunda tüm kümelerin sağlıklı oluşturulmasını garanti edememektedir. Ortalama bağlantı tekniği, bu iki uç teknik arasında sonuçlar vermesi nedeniyle alternatif bir yöntem olarak önerilmektedir (Tatlıdil, 2002).

➢ Küresel Ortalama Bağlantı Kümeleme (Centroid Method):

Bu yaklaşımda ilk olarak her kümenin geometrik merkezi (centroid) hesaplanır. İki küme arasındaki uzaklık bu iki centroid arasındaki uzaklığa eşittir (Mooi ve Sarstedt, 2011).

m kümesinin *j* kümesinden olan uzaklığı;

$$d_{mj} = \frac{N_k d_{kj} + N_l N_{lj}}{N_m} - \frac{N_k N_l d_k}{N_m^2} \tag{3.9}$$

➢ Ward Bağlantı Kümeleme (Ward's Method):

Ward (1963) her grupta bilgi kaybını en aza indirecek şekilde $P_n, P_{n-1}, \dots, P_1$ bölmelerine ayıracak ve bu kaybı nicelleştirecek bir hiyerarşik kümeleme yöntemi önermiştir. Analizin her adımında, birleştiğinde bilgi kaybı en az artışa sahip olacağı düşünülen küme çiftleri birleştirilir. Bilgi kaybı Ward tarafından "hata kareler toplamı kriteri, ESS" olarak tanımlanmıştır (Everitt, 1993).

Başlangıçta her küme tek bir öğeyi oluşturur; ve N öğe var ise $ESS_k = 0, k = 0,1, \dots, N$ olduğundan $ESS_k = 0$'dır. Diğer uçta bütün kümeler N öğeli tek bir grupta toplandığında ise ESS değeri;

$$ESS = \sum_{j=1}^{N} (x_j - \bar{x})'(x_j - \bar{x}) \quad (3.10)$$

x_j ; j. öğeye ilişkin çok değişkenli ölçüm iken $\bar{x}$ ise tüm öğelerin ortalamasıdır (Johnson ve Wichern, 1998).

3.3.1.2. Ayırıcı hiyerarşik kümeleme yöntemleri

Birleştirici hiyerarşik yöntemler her bir birimi tek başına birer küme olarak kabul ederek başlarlarken, ayırıcı hiyerarşik kümeleme yöntemleri tüm birimleri tek bir küme elemanı olarak kabul edip art arda daha küçük kümelere bölerek devam eder (Gore, 2000).

Ayırıcı hiyerarşik kümeleme yöntemleri yukarıdan-aşağıya stratejisini kullanır. Bu strateji tüm birimlerin tek bir kümede yer almasıyla başlar, bu küme hiyerarşinin kökünü oluşturur. Daha sonra kök, kendi içinde daha küçük kümelere, o kümeler de yinelemeli olarak kendinden daha küçük alt kümelere bölünür. Bu bölme işlemi her küme kendi içinde tutarlı en alt kümeye bölünene kadar devam eder (Han vd., 2012).

3.3.2. Hiyerarşik olmayan kümeleme yöntemleri

Birimlerin kendi içinde homojen ve kendi aralarında heterojen olan kümelere ayrılmasını hedefleyen ve prototip kümeler aracılığı ile alt popülasyonların parametre tahminlerini yapmayı (grup ya da küme ortalama vektörleri ve kovaryans matrisleri) amaçlayan yöntemlerdir. Hiyerarşik kümelemede hem birimler hem de değişkenler birbirleriyle değişik benzerlik düzeylerinde kümeler oluştururken, hiyerarşik olmayan yöntemlerde sadece birimler kümelenir. Birimlerin uygun oldukları kümelerde toplanmaları ve n birimin k sayıda kümeye parçalanması hedeflenmektedir (Özdamar, 2010).

Hiyerarşik olmayan yöntemler, değişkenler yerine birimleri, *k* kümede toplayacak şekilde gruplamak için tasarlanmıştır. Küme sayısı, *k*, daha önceden belirlenmiş ya da kümeleme prosedürünün bir parçası olarak tespit edilmiş olabilir. Hiyerarşik olmayan yöntemler birimlerin gruplara başlangıç bölünmelerinden ya da başlangıç çekirdek nokta (seed points) setlerinden başlayabilir. Başlangıç yapılandırmaları yanlılıktan uzak olduğunda iyi bir seçim yapılmış olur. Bir yöntem de, çekirdek noktaların birimler arasından rassal olarak seçilmesi ya da birimlerin rassal olarak başlangıç gruplara bölünmesidir (Johnson ve Wichern, 1998).

Hiyerarşik olmayan yöntemler arasında en yaygın kullanılanları, k-ortalamalar kümeleme (k-means Clustering, MacQueens' Method), Medoid kümeleme (Medoid Clustering) ve Fuzzy (Bulanık) kümeleme (Fuzzy Clustering) yöntemleridir.

3.3.2.1. k-ortalamalar (k-means) kümeleme

MacQueen (1967), her bir birimin en yakın centroid (ortalama) ile k kümeye atanması sürecini ortaya koyan yöntemi, "k-ortalamalar" terimi ile adlandırmıştır. Bu kümeleme yönteminde süreç, centroidlerin son küme üyeliklerine göre değil, o anki üyeliklerinin hesaplanması temeline dayanarak kurulmuştur (Anderberg, 1973).

k-ortalamalar algoritması, geniş çaplı kullanılan bir kümeleme uygulamasıdır. Adını, küme merkezi olarak tanımlanan k tane kümenin her birini temsil eden noktaların ortalamalarından alır. Sayısal nitelikler için iyi bir geometrik ve istatistiksel anlama sahiptir (Erilli, 2009).

Bu centroid tabanlı kümeleme yöntemi kümeyi temsil etmek üzere küme merkezi c_i'yi kullanır. Kavramsal olarak, bir kümenin centroidi o kümenin merkez noktasıdır. Burada centroidin tanımlanmasında birçok farklı yol mevcuttur: kümelere atanan birimlerin ya da noktaların ortalamasının ya da medoidinin alınması gibi (Han vd., 2012).

k-ortalamalar kümeleme yönteminin değerlendirilmesinde en yaygın olarak hata kareler toplamı (Sum of Squared Error-*SSE*) kullanılır. En küçük *SSE* değerine sahip kümeleme, bu kümelemede centroidlerin kümelerin en iyi temsil eden noktalar olduğu anlamına gelir. Hata kareler toplamı (*SSE*) şöyle tanımlanmaktadır:

$$SSE = \sum_{i=1}^{k} \sum_{x \epsilon c_i} dist(c_i, x)^2 \qquad (3.11)$$

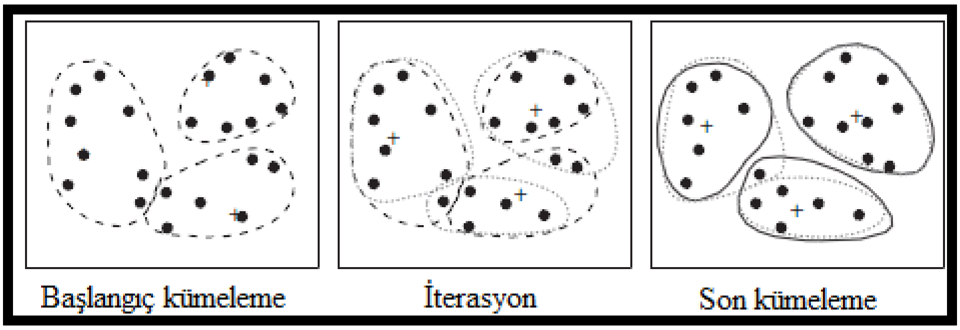
Bu varsayımlara göre *SSE* değerini minimize eden centroid (küme merkezi), kümenin ortalamasıdır. i. kümenin centroidi (ortalaması) aşağıdaki denklemle tanımlanabilir.

$$c_i = \frac{1}{m_i} \sum_{x \in c_i} x \qquad (3.12)$$

m_i : i. kümedeki birim sayısı

(Pang-Ning vd., 2006).

k-ortalamalar kümeleme algoritması şu şekilde çalışmaktadır: Kümelerdeki noktaların ortalama değerini küme merkezi (centroid) olarak tanımlar. Sonra algoritma şu şekilde devam eder. Önce n adet birimden rassal olarak k kadar birim seçer, bu birimlerin her biri küme merkezini temsil etmektedir. Kalan birimlerin her biri için kendisine en benzeyen (en yakın olan) küme merkezine göre atama yapılır. Her küme için bir önceki iterasyon birimleri kullanılarak atama yapar. Bu işlem atamalar sabitleninceye kadar devam eder (Han vd., 2006). Şekil 3.2' de k-ortalamalar kümeleme yöntemi ile birimlerin kümelenmesi aşamaları verilmiştir.



Şekil 3.2. k-ortalamalar kümeleme yöntemi ile birimlerin kümelenmesi, küme merkezlerinin güncellemesi ve buna göre yeniden atanması (her küme ortalaması + ile gösterilmiştir) (Han vd., 2006).

k-ortalamalar yönteminin bazı dezavantajları vardır. Bunlar şöyle sıralanabilir:

i. İlk bölünmeleri ve küme sayısı *k'* yı belirlemede genel olarak etkili kabul edilen bir yöntem yoktur. Centroidlerin yakınsaması farklı başlangıç noktaları ile değişkenlik gösterir. Bu sorunun çözümüne yönelik olan genel yaklaşım rassal başlangıç bölünmeleriyle algoritmayı birçok kez yinelemektir.
ii. k-ortalamaların iteratif optimum prosedürü, global optimuma yakınsaklığı garanti etmez.
iii. k-ortalamalar aykırı değerleri ve gürültüye karşı duyarlıdır. Bir nesne küme merkezinden (centroid) oldukça uzak olsa bile kümeye girmeye zorlanır ve bu küme şeklini bozabilir.
iv. k-ortalamalar tanımındaki "ortalama" ifadesi uygulamayı sadece sayısal verilerle sınırlandırmaktadır (Xu ve Wunsch, 2005).

3.3.2.2. k-medoid kümeleme

k-medoid algoritması, k-ortalamalar yöntemine çok benzemekle beraber, k-ortalamalar yönteminin aykırı değerlere karşı olan duyarlılığını azaltmak amacıyla oluşturulmuştur. Kümelerdeki birimlerin ortalamalarını temsilci olarak almak yerine, kümeler, her küme için birer gerçek birim seçilerek temsil edilebilir. Kalan birimler de bu temsilci birimle olan benzerliğe göre atanır. Bu bölümleme yöntemi her bir birim ile ona karşılık gelen temsilci arasındaki farklılıkların toplamını minimize etmeyi amaçlayan bir yöntemdir (Han vd.,2006).

k-medoid algoritmasının birçok farklı yaklaşımı mevcuttur. Bunlardan en yaygın olarak kullanılanı Kaufman ve Rousseuw (1990) tarafından geliştirilen “Partitioning Around Medoids- *PAM*” algoritmasıdır. *PAM* algoritmasında k adet temsilci “medoid” olarak adlandırılır ve bu medoidler, her bir birimin kendisine en yakın medoidle uzaklıklarının toplam farkını minimum yapacak şekilde hesaplanır. Minimize edilen amaç fonksiyonu şu şekildedir:

$$\sum_{i=1}^{n} \min_{t=1,\ldots,k} d(i, m_t) \tag{3.13}$$

(Struyf vd., 1997)

k-medoid yöntemi, verilerde aykırı değerler bulunsa da iyi sonuçlar verir. Bunun yanında bu algoritmaya girdi değeri olarak k küme sayısının verilmesi gerekmektedir. Bu nedenle iyi bir kümeleme elde etmek için k sayısının ne olacağına karar vermek gerekir. Bu değerin kullanıcıya bırakılması önemli bir dezavantajdır (Erilli, 2009).

Uygun kümelemede çekirdek (medoid) sayısı ve bu çekirdek noktalarına göre belirlenen kümelerin uygunluğu için *gölge (siluet) istatistiği*’ nden yararlanılır. Gölge istatistiği (s) şu şekilde hesaplanır:

i. *A* kümesindeki *n* birimden *i*. birimin tüm diğer birimlere olan uzaklıkları ortalaması *a* belirlenir.

$$a = \frac{1}{n}\sum_{j=1}^{n} d_{ij} \; ; n \in A \tag{3.14}$$

Eğer kümedeki birim sayısı $n=1$ ise $a=0$ alınır.

ii. *A* kümesi dışında fakat *i*. birimin en yakın komşu olduğu ve elemanları arasındaki ortalama farklılığın en küçük olduğu *B* kümesindeki elemanlar ile *i*. birimin uzaklıklarının ortalaması *b* belirlenir.

$$b = \frac{1}{n}\sum_{j=1}^{n} d_{ij} \; ; n \in B, i \in A \tag{3.15}$$

iii. *a* ve ortalama değerleri kullanılarak *i*. birimin gölge istatistiği *s* aşağıdaki kurallara göre hesaplanır.

Eğer *A* kümesi eleman sayısı $n=1$ ise $s=0$
Eğer $a<b$ ise $s=1-a/b$
Eğer $a>b$ ise $s=b/a-1$
Eğer $a=b$ ise $s=0$ olarak alınır.

Tüm birimler için gölge istatistiği *s* , *+1* ile *-1* arasında değer gösterir. *s, +1* ‘e yakın ise *i*. birim doğru sınıflanmıştır, sıfıra yakınsa *i*. birim *A* ve *B* kümeleri arasında olup *A* kümesine atandığı varsayılır, *-1*’e yakınsa *i*. birim A kümesine yanlış atanmıştır.

Kaufman ve Rousseuw medoid kümeleme yönteminde küme sayısını belirlemek için s değerleri ortalaması (ortalama gölge istatistiği, *SC*) istatistiğinden yararlanılır. *SC*’nin en büyüklendiği çözüm en uygun çözüm olarak alınır (Özdamar, 2010).

4. BULANIK (FUZZY) KÜMELEME

Bulanık kümelemeye geçilmeden önce bulanık mantık kavramı ele alınacaktır.

4.1. Bulanık Mantık

Bilimsel çalışmalar ve problem çözümlemede kullanılan sistemler gerçekliğin matematiksel modellemesi temeline dayanmaktadır. Bu matematiksel modeller, gerçekteki olayın ya da durumun karmaşıklığı, güvenilirliği, belirsizliği ve bunların arasındaki ilişkiyi ortaya koymayı amaçlar. Oluşturulan model, ne kadar ayrıntılı olursa olsun; gerçek dünyanın çoğu yönden belirsiz, karmaşık ve kesinlikten yoksun oluşundan dolayı, tam anlamıyla gerçeği yansıtamamaktadır.

Klasik mantık, bilindiği üzere önermelerin doğru ya da yanlış oluşuyla ilgilenir. Her önermenin bir zıttı vardır. Klasik mantık bu önermelere ait değişkenlerin kombinasyonları ile ifade edilir. Her değişken hipotetik olarak, herhangi bir kombinasyonun sonucunda gerçek bir değere sahip olduğu varsayılan (doğru ya da yanlış) bir önermeye dayandırılır ama bu değer asla ikisinin arasında yer almaz (aynı anda doğru ve yanlış olamaz) (Chen ve Pham, 2001). Başka bir deyişle; ikili mantıkta değişkenlerin alabileceği birbirinin zıttı olan iki farklı değer vardır ve ancak ve ancak değişkenler bu iki zıt değerden (doğru-yanlış, evet-hayır, sıcak-soğuk, vb.) birini alabilir.

Bulanık mantığın ardındaki temel fikir, bir önermenin ‘doğru’, ‘yanlış’, ‘çok doğru’, ‘çok yanlış’, ‘çok çok doğru’, ‘çok çok yanlış’, ‘yaklaşık olarak doğru’, ‘yaklaşık olarak yanlış’, vb. gibi, olabileceğidir. Diğer bir deyişle, doğruluk, önermelerle klasik yanlış ve doğru arasındaki sonsuz sayıdaki doğruluk değerlerini içeren bir kümedeki değerler, ya da sayısal olarak **[0, 1]** gerçel sayı aralığıyla ilişkilendiren bir fonksiyondur (Ünler, 2006).

Bulanık mantığın ve bu mantık kurallarını kullanan bulanık küme teorisinin Lotfi A. Zadeh tarafından geliştirilip 1965 tarihli orijinal makalesinde (Zadeh,1965) yayınlanmasından sonra belirsizlik içeren sistemlerin incelenmesi yeni bir boyut kazanmıştır. 1965'de ortaya atılmasına rağmen, bulanık küme kavramı ancak 1970'li yılların ikinci yarısından sonra kullanılmaya başlanmıştır (Altaş, 1999).

Bulanık mantık, klasik mantığın yadsıdığı belirsizliğin matematiksel olarak ifade edilmesine olanak sağlar. Her ne kadar pozitif bilimler belirsizlikle değil, kesin olanla ilgilense de gerçek yaşam sadece kesinliklerden oluşmaz. Bulanık mantık, insan doğasından yola çıkarak, insana ait dilsel değişkenler yardımıyla belirsizlikleri ifade etmeye çalışır.

Bulanık mantığın birçok uygulama alanı vardır. 1970'li yıllardan itibaren, Zadeh (1965) tarafından ortaya atılan ve daha sonra Mamdani ve Assilian (1973; 1974) tarafından geliştirilen yöntemlerle bulanık mantık kontrol sistemlerinde çokça uygulanır hale gelmiştir. İlk uygulaması 1974 yılında Mamdani tarafından bir buhar makinesinde olup, ticari olarak ise ilk defa Danimarka'da bir çimento fabrikasında kullanılmıştır. Dünyanın en gelişmiş metrosu kabul edilen Sendai metrosu da bu sistemle çalışmaktadır. Kontrol sistemlerinin yanı sıra bulanık mantık; bilgi sistemleri, görüntü tanımlama, biyoinformatik, veri madenciliği, jeoloji, işletme ve yöneylem araştırması gibi çok çeşitli alanlarda kullanılmaktadır.

Bulanık mantık, Zadeh (1965)' in ortaya attığı bulanık küme teorisinin genişletilmiş halidir. İzleyen bölümde bulanık küme kavramına değinilecektir.

4.1.1. Bulanık küme teorisi

Klasik küme teorisi üye olma, üye olmama kavramları üzerine kurulmuştur. Klasik küme teorisinde, herhangi bir kümenin elemanın o kümeye ait olup olmaması ayrımında keskin, kesin, belirli bir sınır vardır. Diğer bir deyişle "Bu eleman küme ait midir?" sorusunun cevabı ancak "evet" ya da "hayır" olabilir. Bu deterministik ve stokastik durumlarda doğru olabilir. Ancak, olasılıkta ve istatistikte “ Bu elemanın kümeye ait olma olasılığı nedir?” şeklinde bir soru sorulabilir. Bu durumda cevap, ” Bu elemanın kümeye ait olma olasılığı %90’dır.” şeklinde olsa bile o eleman hem kümenin üyesidir hem de değildir. Doğru bir tahmin yapılmış olması durumunda eleman %90 kümeye aittir ama bu aynı zamanda %10 kümeye ait olmadığı anlamına gelir. Klasik küme teorisinde bir elemanın aynı anda üye olma ve üye olmamasına izin verilmez. Bu yüzden gerçek dünyadaki birçok problemin elemanları kısmi üyeliklere sahip olduğundan, klasik küme teorisiyle açıklanamaz ve ele alınamaz. Aksine bulanık küme teorisi, kısmi üyelikleri kabul eder ve bir anlamda klasik küme teorisini genelleştirir (Chen ve Pham, 2001).

Klasik kümelemede birimler o kümeye ya aittir ya da değildir. Dolayısıyla alabileceği üyelik değerleri yalnızca “0” ve “1”dir. Buna karşılık bulanık kümelemede, birimler kümelere derecelenerek üyelik gösterirler. Örneğin, bir birey “uzun insanlar” kümesine ait bir bireyken, aynı zamanda aldığı üyelik derecesi, bulanık kümede tanımlanan “uzun” kavramına ne kadar uygun olduğunu da belirtir. Bu noktada bir birimin o kümeye ne kadar ait olup ne kadar olmadığını tanımlayacak bir fonksiyon gereklidir. Bu fonksiyona, “üyelik fonksiyonu”, bu fonksiyonun oluşturduğu kümeye de “ bulanık küme” adı verilmektedir. Bulanık kümenin tanımı ise aşağıdaki gibidir:

$X = \{x_1, x_{2,} \dots, x_n\}$ sonlu bir küme olsun. *X*’ in bulanık altkümesi;

$$A = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \cdots + \frac{\mu_A(x_n)}{x_n} = \sum_{i=1}^{n} \mu_A(i)/x_i \tag{4.1}$$

Eğer *X* sonlu bir küme değilse, buna ait bulanık küme;

$$A = \int_X \mu_A(x)/x \tag{4.2}$$

(Klir ve Folger, 1988)

Bulanık kümeler bazı temel özelliklere sahiptir. Bu özelliklere göre bulanık kümeler üç grupta incelenir: normal bulanık kümeler, normal olmayan bulanık kümeler ve konveks bulanık kümeler.

*Normal bulanık küme*de; genellikle üyelik derecesi *1*'e eşit olan en az bir elemanın bulunması gerekmektedir.

Normal olmayan (dış bükey) bulanık küme; elemanlarını tümünün üyelik derecesi 1'den küçük olan kümelerdir.

$x < y < z$ bağıntısını sağlayan A bulanık kümesinin *x, y, z* elemanları için $\mu_A(y) \geq [\mu_A(x), \mu_A(z)]$ oluyorsa, yani; başka bir deyişle *A*'nın artan değerleri için üyelik değerleri monoton artan veya monoton artıp sonra monoton azalan oluyorsa; A kümesi *konveks bulanık küme* adını alır.

Konveks ve normal olan bir bulanık küme, *bulanık sayı* olarak adlandırılır (Karanfil, 1997; Şentürk, 2006).

4.1.2. Bulanık mantık üyelik fonksiyonları

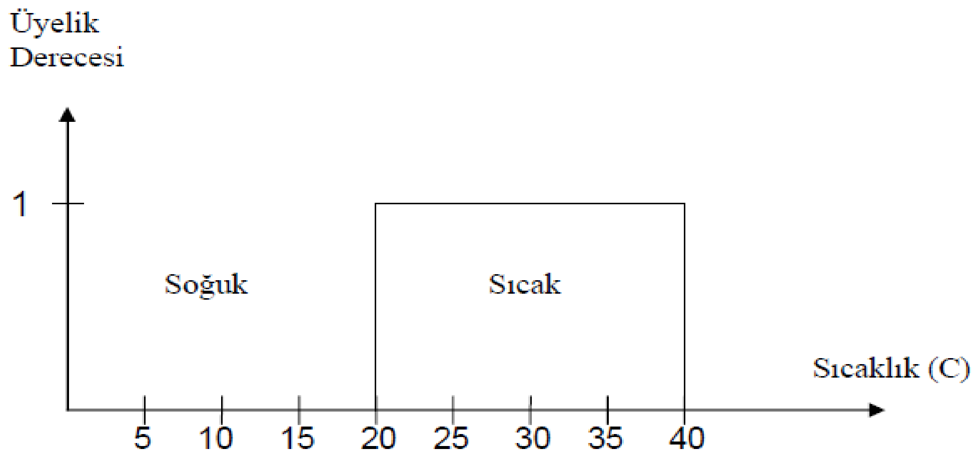
X boş olmayan bir küme ve A, X evrensel kümesinde tanımlı bir bulanık küme olmak üzere,

$$\forall x \in X \text{ } için; \mu_A(x): X \rightarrow [\mathbf{0}, \mathbf{1}] \tag{4.3}$$

Burada, $\mu_A(x)$ bulanık kümeye karşılık gelen üyelik fonksiyonudur. Üyelik fonksiyonu A kümesine ait elemanların aranılan özelliğe ne kadar uygun olup olmadığını ifade eder (Klir ve Yuan, 1997; Zadeh, 1965).

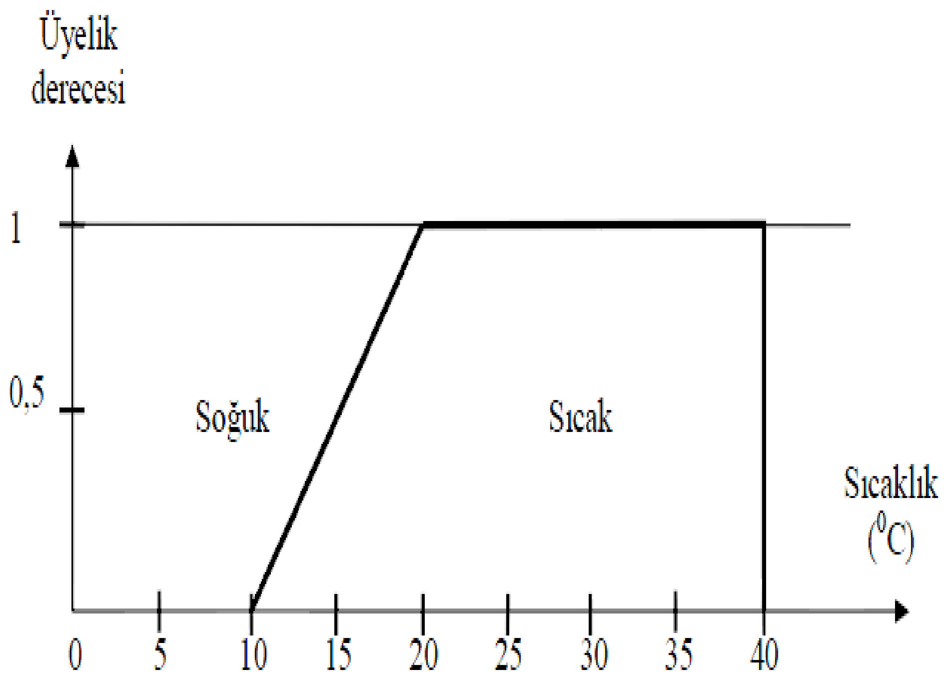
Bulanık mantık, günlük hayatta kullandığımız dilsel değişkenlerden yola çıkarak matematiksel bir ifade ortaya koyar. Bu ifadeyi de bizim dilsel değişkenlerle (çok kısa-kısa-orta-uzun-çok uzun ya da çok az-az-biraz-fazla-çok fazla vb.) nitelendirdiğimiz derecelendirmeyi üyelik fonksiyonu yardımı ile yapar. Burada söz konusu elemanın o kümeye ait olmasının ya da ait olmamasının bilinmesi yetersizdir. Üyelik fonksiyonu yardımıyla söz konusu elemanın o kümeye ne kadar ait olup ne kadar olmadığı ortaya konulur.

Örneğin, klasik küme teorisi altında hava sıcaklığını ele alalım. Klasik küme için hava sıcaklığı örneği olan Şekil 4.1 incelendiğinde, sıcaklık kümesi 20^0C'nin altında soğuk, 20^0C'nin üzerinde ise sıcak olarak tanımlanmıştır.

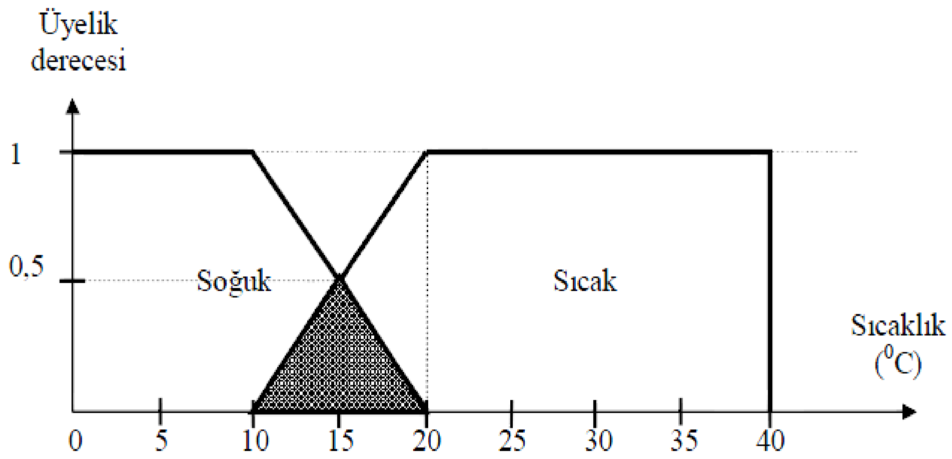


Şekil 4. 1. Klasik küme için hava sıcaklığı örneği

Görüleceği üzere, bu klasik küme örneğinde hava sıcaklığı soğuk-sıcak olmak üzere iki ayrı keskin gruba ayrılmıştır. Oysa gerçek yaşamda böyle bir keskinlikten (sıcak-soğuk, kısa-uzun, genç-yaşlı, aydınlık-karanlık vb.) söz etmek güçtür. Doğadaki kavramlara ait sınırlar bu kadar keskin çizgilerle nitelendirilemez. Bulanık kümelerde ise gerçek yaşamda kullandığımız dilsel değişkenlerle (sıcak-ılık-az soğuk-soğuk, kısa-biraz kısa-orta-biraz uzun, aydınlık-biraz karanlık vb.) bu keskin çizgiler yumuşatılır. Hava sıcaklığı, Şekil 4.2' de bulanık kümeler ile örneklendirilmiştir. Bkz. Şekil 4.2 incelendiğinde 10-40^{0}C arası değerlerin sıcak kümesine üye olduğu, ancak, 20-40^{0}C arasında yer alan değerlerin sıcak kümesine üyeliğinin 1 olduğu görülmektedir. 10-20 ^{0}C arasında yer alan değerler ise sıcak kümesine 0-1 arasında değişen üyelik dereceleriyle ait olacaktır. Şekil göz önüne alınarak, 13^{0}C az sıcak olarak, 17^{0}C ise biraz sıcak olarak; 35^{0}C ise çok sıcak olarak nitelendirilebilir. Burada görüleceği üzere 10-20^{0}C aralığında yer alan değerler hem soğuk hem de sıcak kümesine aittir. Yani soğuk ve sıcak kümesi kesişmektedir. Bu durum bulanık kümelerin ötüşümü olarak adlandırılır. Şekil 4. 3.'de yer alan taralı alan ile örtüşüm gösterilmiştir.



Şekil 4. 2. Bulanık küme için hava sıcaklığı örneği



Şekil 4. 3. Bulanık kümelerde örtüşüm

Bir bulanık alt kümenin üyelik fonksiyonunu seçmek için genelleştirilmiş tek, kesin bir kural ya da kriter yoktur. Doğru ve iyi bir üyelik fonksiyonunun tanımlanması, kullanıcının bilimsel birikimine, iş tecrübesine ve söz konusu uygulama için gerçekten neye ihtiyacı olduğuna dayalı olarak yapılır (Chen ve Pham, 2001).

Bulanık kümeler üyelik fonksiyonu ile karakterize edilirler. Üyelik fonksiyonunun tanımlanması ve için en uygun ve kısa yöntem matematiksel olarak formüle edilmesidir. Bunun için en sıklıkla kullanılan fonksiyonlar; üçgen üyelik fonksiyonu, yamuk üyelik fonksiyonu, gauss üyelik fonksiyonu ve genelleştirilmiş bell üyelik fonksiyonudur (Jang vd., 1997).

4.1.2.1. <u>Üçgen üyelik fonksiyonu</u>

Üçgen üyelik fonksiyonu $\{a, b, c\}$ üç parametre ile özelleştirilmiştir.

$$\text{üç}gen(x; a, b, c) = \begin{cases} 0 \text{ ,} x \leq a. \\ \frac{x-a}{b-a} \text{ ,} a \leq x \leq b. \\ \frac{c-x}{c-b} \text{ ,} b \leq x \leq c. \\ 0 \text{ ,} c \leq x. \end{cases} \tag{4.4}$$

(Jang vd., 1997)

4.1.2.2. Yamuk üyelik fonksiyonu

Yamuk üyelik fonksiyonu $\{a, b, c, d\}$ şeklinde dört adet parametreyle özelleştirilmiştir.

$$yamuk(x; a, b, c, d) = \begin{cases} 0, & x \le a. \\ \frac{x-a}{b-a}, & a \le x \le b. \\ 1, & b \le x \le c. \\ \frac{d-x}{d-c}, & c \le x \le d. \\ 0, & d \le x. \end{cases} \tag{4.5}$$

(Jang vd., 1997, Klir ve Yuan, 1997)

4.1.2.3. Gauss üyelik fonksiyonu

Gauss üyelik fonksiyonu $\{c, \sigma\}$ parametreleriyle özelleştirilmiştir.

$$gauss(x; c, \sigma) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \tag{4.6}$$

c burada üyelik fonksiyonunun merkezini, σ ise genişliğini belirtir (Jang vd., 1997).

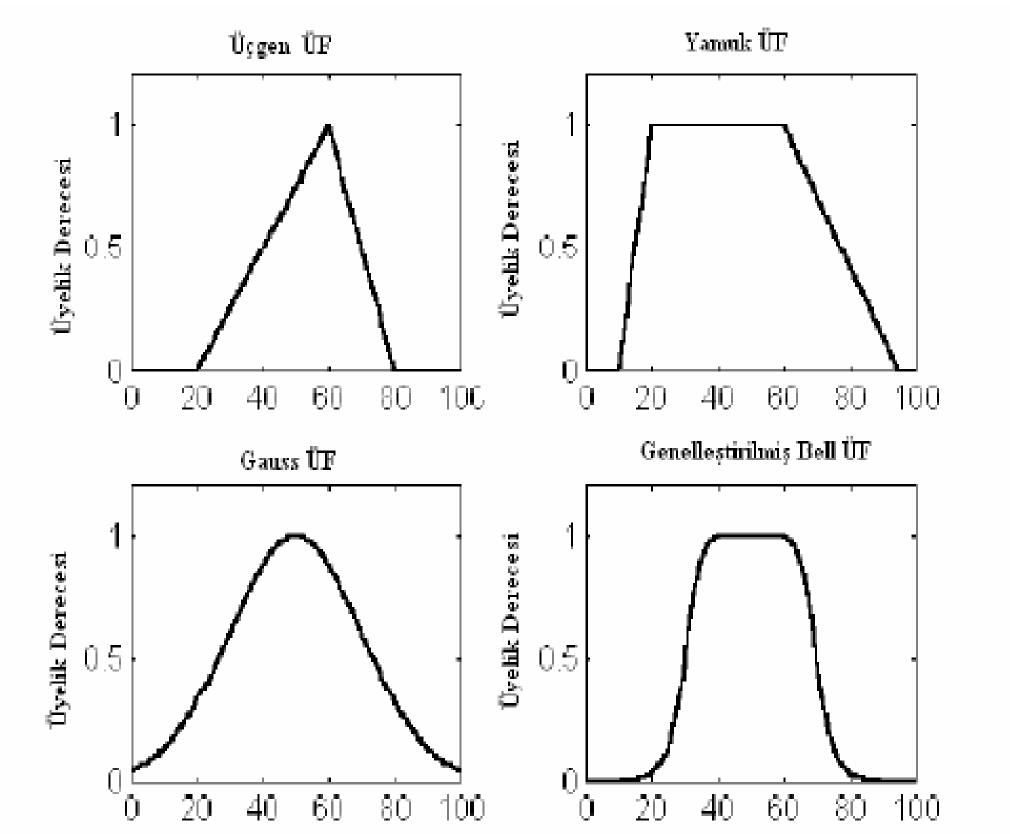
4.1.2.4. Genelleştirilmiş bell üyelik fonksiyonu

Genelleştirilmiş bell üyelik fonksiyonu $\{a, b, c\}$ olmak üzere üç parametre ile özelleştirilmiştir.

$$bell(x; a, b, c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \tag{4.7}$$

b parametresi genel olarak pozitiftir (*b* negatif olduğunda bell üyelik fonksiyonunun şekli aşağıya doğru olur.) Burada *c* üyelik fonksiyonunun merkezini, *a* ise genişliğini belirtir. *b* ise eğimi ve geçiş noktalarını kontrol eder.

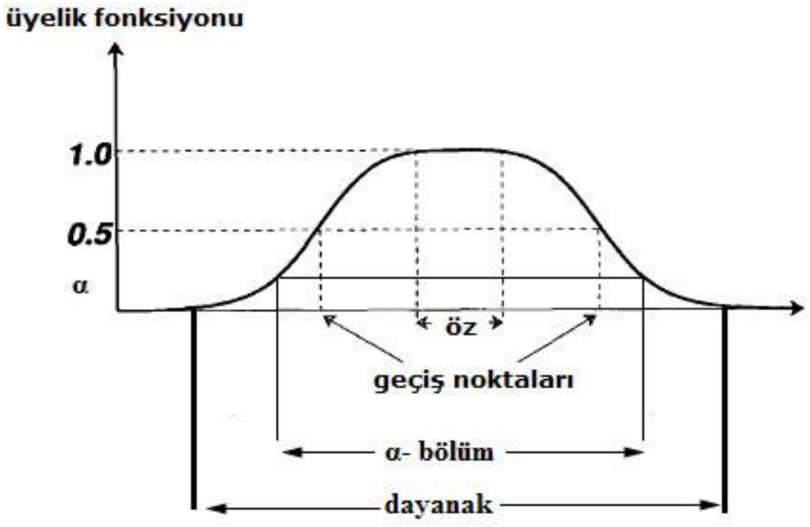
Üçgen, yamuk, gauss ve genelleştirilmiş bell üyelik fonksiyonları Şekil 4. 4.'de gösterildiği gibidir (Jang vd., 1997).



Şekil 4. 4. Bazı üyelik fonksiyonları

4.1.3. Bulanık mantık üyelik fonksiyonlarının kısımları

Üyelik fonksiyonları genel olarak öz (core), geçiş noktaları, α-bölüm (cut) ve dayanak (support) olmak üzere dört kısma ayrılır (Elmas, 2003; Karanfil, 1997). Üyelik fonksiyonunun kısımları Şekil 4.5' te verilmiştir.



Şekil 4. 5. Üyelik fonksiyonunun kısımları

A bulanık kümesine tam üyeliğe sahip elemanların oluşturduğu bölgeye üyelik fonksiyonunun *öz*ü denir (Karanfil, 1997). Başka bir deyişle üyelik derecesi *1*'e eşit olan elemanların oluşturduğu bölge *A* bulanık kümesinin özü (çekirdek, core) olarak tanımlanır.

Üyelik dereceleri *0.5*' e eşit olan elemanların toplandığı *A* altküme kısmı o alt kümenin *geçiş noktaları* olarak adlandırılır (Şentürk, 2006).

Bulanık *A* kümesinin α-bölüm kümesi A_a ile gösterilir ve *X* evrensel kümesinin *A* kümesindeki bütün elemanlarının üyelik derecesi α özel değerinden büyük veya eşit olanları içerir (Elmas, 2003).

A bulanık kümesinde sıfır olmayan üyeliğe sahip elamanların oluşturduğu bölgeye *A* bulanık kümesinin dayanağı (support) denir (Karanfil, 1997).

4.2. Bulanık (Fuzzy) Kümeleme

Kümeleme yöntemleri; verinin kümelere nasıl atandıklarına, başka bir deyişle, hangi türde bölünmeler oluşturduklarına göre ayrılırlar. Klasik kümeleme yöntemlerinde her birim kesin olarak bir kümeye atanmak zorundadır. Bu yöntemler, veri setini eksiksiz olarak boş olmayan ve ikili ayrık alt gruba ayrıştıran bölünmeler üretir. Veriyi kümeye böyle katı bir şekilde atamak, iki ya da daha fazla kümeye eşit mesafede olan veri noktalarının(birimlerin) varlığında yetersiz kalabilir. Bu katı bölümleme birim(ler) aynı anda iki ya da daha fazla kümeye eşit oranda üyeyken, onları sadece tek bir kümeye ait olmaya zorlar. Kümelerin anlamlılığını geliştirmek üzere oluşturulan yöntemler, böyle kesişen kümeler durumunda, bu "keskin" kümeleme yönteminin birim(ler)in aynı anda iki kümeye de atanmasına olanak vermektedir. Buna rağmen, klasik kümeleme yöntemleri birimin farklı kümelere atanırken, ne kadar kesin, ne kadar kesin olmayan bir şekilde atandığını göstermez (Döring vd., 2006).

Zadeh (1965) tarafından geliştirilen bulanık küme teorisi ile açıklanan "üyelik fonksiyonu" bu ait olma durumunun belirsizliği hakkında bir fikir ortaya koymuştur. Bulanık kümelerin bu kullanımı, belirsiz küme üyelikleri hakkında bilgi sağlar. Bulanık küme teorisinin kümeleme analizinde uygulamaları ilk olarak, Bellman, Kalaba, Zadeh (1966) ve Ruspini (1969) tarafından yapılan çalışmalarda ileri sürülmüştür. Bu çalışmalar bulanık kümeleme araştırmalarının kapısı açmıştır (Yang, 1993).

Bulanık küme teorisiyle kümelemenin bir sentezi olan bulanık kümeleme, belirsiz küme sınırlarına sahip problemleri ele almak için uygun bir yöntemdir. Bulanık kümeleme, her bir birimin sadece tek bir kümeye atanma zorunluğunu gevşeterek, her bir birimin belli üyelik dereceleriyle kümelere üye olduğu daha zayıf bir zorunluluk haline getirir. Ayrıca, bu üyelikler birimler ve kümeler arasında daha karmaşık olan ilişkilerin ortaya çıkmasına yardım eder (Mansoori, 2011).

Diğer kümeleme yöntemleri gibi, bulanık kümeleme yöntemi de uzaklık ölçüsü temellidir. Uygulamalarda Öklid ve Mahalanobis uzaklıklarının yanı sıra, diğer uzaklık ölçülerinin de kullanıldığı birçok farklı bulanık kümeleme algoritması mevcuttur. Bulanık kümelemenin birçok kullanışlı özelliği vardır. Bunlardan bazıları;

i. Yorum açısından faydalı olan üyelik değerlerini sağlar.
ii. Uzaklık kullanımı konusunda esnektir.
iii. Üyelik değerlerinden bazıları biliniyorsa, bu sayısal optimizasyon ile birleştirilebilir (Naes ve Mevik, 1999).

Bulanık kümelemede iki temel yöntem vardır (Klir ve Yuan,1997). Bunlardan ilki bulanık kümeleme içi ilişkileri, diğeri ise amaç fonksiyonunu kullanır. Bulanık ilişkileri temel alan kümeleme yöntemi birimler arası ilişkisel yapıyla ilgilenir. Amaç fonksiyonunu temel alan kümeleme yöntemi ise, amaç fonksiyonu ile kümeleme problemini optimizasyon problemi haline dönüştürerek çözümleme yapar.

En sık kullanılan bulanık kümeleme algoritmaları aşağıdaki gibi sınıflandırılabilir:

a) Geleneksel Bulanık Kümeleme Algoritmaları
   i. Bulanık c- ortalamalar
   ii. Gustafson-Kessel
   iii. Gath-Geva

b) Prototipi Farklı Geometrik Şekle Sahip Kümeleme Algoritmaları
   i. Bulanık c-regresyon
   ii. Bulanık c-hatlar
   iii. Uyarlamalı Bulanık Kümeleme
   iv. Kabuk Prototip

Çalışmada yer alan uygulamada kullanılan geleneksel bulanık kümeleme algoritmaları izleyen bölümde açıklanacaktır.

4.2.1. Geleneksel bulanık kümeleme algoritmaları

Çoğu bulanık kümeleme algoritması c-ortalamalar amaç fonksiyonunun optimizasyonuna ve bu fonksiyonun modifikasyonlarına dayanır. Amaç fonksiyonuna dayalı kümelemede, her küme bir küme prototipi tarafından temsil edilir. Bu prototip kümenin merkezi, şekli ve boyutu hakkında bilgiler içerir. Küme merkezi etki alanındaki kümenin özniteliğini temsil etmek ve aynı zamanda veri noktalarını bölmek için kullanılır. Ancak, kümeleme algoritması tarafından hesaplanan küme merkezi, veri setinde bulunuyor ya da bulunmuyor olabilir. Şekil ve ölçek parametreleri etki alanındaki kümenin altında yatan farklı yönlerdeki uzantıları belirler (Döring vd., 2006).

Birimlerin hangi kümeye ait olduklarını gösteren üyelik fonksiyonları birimlerle küme merkezlerinin birbirlerine olan uzaklıkları hesaplanarak bulunur. Kümelerin mümkün olduğunca homojen olması gerektiğinden, burada sorun veriyi küme merkezleriyle birimler arası uzaklıklar toplamını minimum yapacak şekilde *c* kümeye atamaktır (Döring vd., 2006).

4.2.1.1. Bulanık-c ortalamalar (fuzzy c-means- FCM) algoritması

Amaç fonksiyonuna dayalı kümeleme yöntemleri arasında en sıklıkla kullanılan algoritmalardan biri bulanık-c ortalamalar algoritmasıdır. Bulanık-c algoritmasının ilk formu Dunn (1973) tarafından sunulmuş, daha sonra Bezdek (1981) tarafından tamamlanmıştır (Grekousis ve Thomas, 2012). Çoğu bulanık kümeleme algoritması, Dunn ve Bezdek tarafından formüle edilen bulanık c-ortalamalar fonksiyonunun minimize eden bu fonksiyon temel alınarak oluşturulmuştur.

Bulanık c-ortalamalar metodu, nesnelerin (birimlerin) iki veya daha fazla kümeye ait olabilmesine izin verir. Bulanık mantık prensibi gereği her veri, kümelerin her birine **[0,1]** arasında değişen birer üyelik değeri ile aittir. Bir verinin tüm sınıflara olan üyelik değerleri toplamı *"1"* olmalıdır. Nesne hangi küme merkezine yakın ise o kümeye ait olma üyeliği diğer kümelere ait olma üyeliğinden daha büyük olacaktır (Işık ve Çamurcu, 2007).

FCM algoritması *p* boyutlu veri setini, bilinen ya da verilen *c* sayıdaki küresel noktalar kümelerine ayırır. Kümelerin yaklaşık olarak aynı boyutlu olduğu kabul edilir. Her küme, kümenin merkezi olan prototipler tarafından temsil edilir. Birimler ve küme merkezleri arasındaki uzaklık, Öklid uzaklık ölçüsü ile belirlenir (Höppner vd., 1999).

FCM algoritmasının amaç fonksiyonu şu şekildedir:

$$J(X;U,V) = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m \left\| x_j - v_i \right\|^2 \quad (4.8)$$

Burada küme merkezi (prototip) vektörü;

$$V = [v_1, v_2, \ldots, v_c] \quad (4.9)$$

iken; küme merkezleri (prototipler) ve birimler arası Öklid uzaklığı ise;

$$D_{ij}^2 = \left\| x_j - v_i \right\|^2 = \left(x_j - v_i\right)^T (x_j - v_i) \quad (4.10)$$

şeklinde hesaplanır.

Fonksiyonda *c* küme sayısını, *n* birim sayısını göstermektedir. μ_{ij} , *j*. kümedeki x_i üyeliğini göstermekte ve $0 \leq \mu_{ij} \leq 1$, $0 < \sum_{j=1}^{n} \mu_{ij} < n$, $\sum_{i=1}^{c} \mu_{ij} = 1$ şeklindedir. $m \geq 1$ olup, bulanıklığı ağırlıklandırma katsayısıdır.

Amaç fonksiyonunu minimize etmek için iki farklı koşul vardır (Yang, 1993);

$$v_i = \frac{\sum_{j=1}^{n} \left(\mu_{ij}\right)^m x_j}{\sum_{j=1}^{n} \left(\mu_{ij}\right)^m} \ , 1 \leq i \leq c \quad (4.11)$$

$$\mu_{ij} = \left(\sum_{k=1}^{c} \frac{\|x_j - v_i\|^{2/(m-1)}}{\|x_j - v_k\|^{2/(m-1)}} \right)^{-1}, i = 1, \ldots, c \; ve \; j = 1, \ldots, n \qquad (4.12)$$

FCM algoritması ayrıntılı olarak şu şekilde işlemektedir:

Adım 1: c küme sayısı, m bulanıklık katsayısı, i iterasyon sayısı, ε işlem sonlanma kriteri olmak üzere başlangıç değerleri belirlenir; μ_{ij} ‘lerden oluşan U üyelik matrisine rasgele değerler atanır.

Adım 2: (4.11) eşitliği kullanılarak küme merkezleri (prototipler) hesaplanır.

Adım 3: (4.10) eşitliği ile küme merkezleri ve birimler arası Öklid uzaklığı hesaplanır.

Adım 4: Yeni üyelik değerleri (4.12) eşitliği ile hesaplanır.

Adım 5: Yeni üyelik değerleri ile eski üyelik değerleri karşılaştırılır.
$\|U_{yeni} - U_{eski}\| \leq \varepsilon$ ise iterasyon durdurulur. Aksi takdirde Adım 2’ye geri dönülür (Çemrek vd.,2010; Türkşen, 2005; Avcı, 2006)

FCM algoritması hangi birimin hangi kümeye atanacağına üyelik derecelerini kullanarak karar verir. Her bir birim, hangi kümeye olan üyeliği büyükse o kümeye atanır. Ancak bunun yanında, her bir birim diğer kümelere de belli üyelik dereceleriyle üye olabilir.

FCM algoritması başlangıçta rasgele atanan değerlere bağlıdır ve bu değerleri kullanarak iteratif olarak güncelleme yapar. Algoritma sonunda ise küme merkezleri veri seti içinde olması gereken yere konumlar.

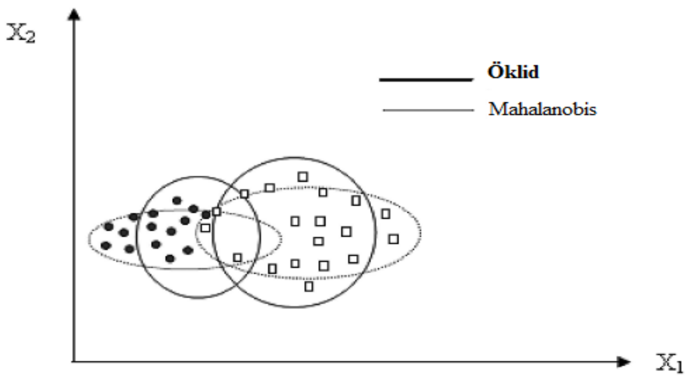
Küme merkezlerinin ilk yerleri, başlangıçta değeri rasgele atanan *U* matrisi kullanılarak oluşturulduğu için, bulanık c-ortalamalar optimal sonuca yaklaşmayı garanti etmeyecektir (Sintas vd., 1999).

Performans; merkezlerin başlangıç yerlerine bağlıdır. Daha güçlü bir yaklaşım tanımlanan iki yol vardır:

i. Tüm merkezleri tanımlamak için bir algoritma kullanmak.
ii. *FCM*’ yi farklı başlangıç merkezleri ile tekrarlı olarak çalıştırmak (Avcı, 2006).

4.2.1.2. Gustafson- Kessel (GK) algoritması

Gustafson-Kessel algoritması elips şeklindeki kümeleri tespit etmek amacıyla Gustafson ve Kessel (1979) tarafından oluşturulan, bulanık kümeleme algoritmasının uyarlanmış bir halidir. Gustafson-Kessel algoritması, Öklid uzaklığı yerine Mahalanobis uzaklığını kullanır. Öklid uzaklığı küresel kümeler oluştururken, Mahalanobis uzaklığı ise elips şeklinde kümeler oluşturur. Şekil 4.6’ da Öklid ve Mahalanobis uzaklık ölçüleri uygulandığında kümelemenin küresel ve eliptik olarak değişimi gösterilmiştir.



Şekil 4. 6. Öklid ve Mahalanobis uzaklık ölçüleri uygulandığında kümelemenin küresel ve eliptik olarak değişimi (Türkşen, 2005)

Gustafson-Kessel algoritmasında, her küme simetrik ve pozitif tanımlı bir *A* matrisine sahiptir. *A* matrisi uzaklık ölçüsünde, verideki topolojik yapıya göre her bir küme için normu hesaplamak amacıyla kullanılır (Grekousis ve Thomas, 2012).

Gustafson-Kessel algoritmasının amaç fonksiyonu;

$$J(X;U,V,A) = \sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^m D_{ijAi}^2 \tag{4.13}$$

şeklindedir.

Bu fonksiyonda D_{ijAi}^2 Mahalonobis uzaklığı;

$$D_{ijAi}^2 = \left(x_j - v_i\right)^T A_i\left(x_j - v_i\right),\ 1 \leq i \leq c, 1 \leq i \leq n \tag{4.14}$$

olarak hesaplanır. A norm matrisi, $A = (A_1, A_2, \dots, A_c)$ ise;

$$\|A_i\| = \rho_i,\ \rho > 0 \tag{4.15}$$

$$\|A_i\| = [\rho_i \det(F_i)]^{1/n} F_i^{-1} \tag{4.16}$$

denklemleriyle hesaplanır. F_i , kovaryans matrisi S_i yerine kullanılan *bulanık kovaryans matrisi*dir. *i*. kümenin bulanık kovaryans matrisi aşağıdaki eşitliğe göre hesaplanır (Balasko vd., 2005).

$$F_i = \frac{\sum_{j=1}^{n}(\mu_{ij})^m (x_j - v_i)(x_j - v_i)^T}{\sum_{j=1}^{n}(\mu_{ij})^m} \tag{4.17}$$

Amaç fonksiyonu minimize edildiğinde bulanık küme merkezleri aşağıdaki gibidir.

$$v_i = \frac{\sum_{j=1}^{n}(\mu_{ij})^m x_j}{\sum_{j=1}^{n}(\mu_{ij})^m}\ , 1 \leq i \leq c \qquad (4.18)$$

Gustafson-Kessel algoritmasının izlediği adımlar şu şekildedir:

Adım 1: c küme sayısı, m bulanıklık katsayısı, i iterasyon sayısı, ε işlem sonlanma kriteri olmak üzere başlangıç değerleri belirlenir; μ_{ij} 'lerden oluşan U üyelik matrisine rasgele değerler atanır.

Adım 2: Eşitlik (4.18)' den yararlanılarak bulanık küme merkezleri hesaplanır.

Adım 3: Eşitlik (4.17) kullanılarak her küme için bulanık kovaryans matrisi hesaplanır.

Adım 4: (4.14) eşitliği kullanılarak her bir birim için Mahalanobis uzaklığı hesaplanır.

Adım 5: (4.12) eşitliği ile yeni üyelik değerleri matrisi hesaplanır.

Adım 6: $\|U_{yeni} - U_{eski}\| \leq \varepsilon$ ise iterasyon sona erdirilir; aksi takdirde Adım 2 'ye geri dönülür (Balasko vd., 2005).

4.2.1.3. Gath-Geva (GG) algoritması

Gath-Geva algoritması, Gath ve Geva (1989) tarafından maksimum olabilirlik tahmincisinin bulanık kümeleme için genelleştirilmiş halidir. Gath-Geva algoritması farklı şekil, boyut ve yoğunluktaki kümeleri tespit etmek amacıyla kullanılır (Abonyi ve Feil, 2007). Bulanık maksimum olabilirlik tahmini, bulanık maksimum tahmincisi uzaklık normunu kullanır. Bu norm, Gustafson- Kessel algoritmasına nazaran uzaklığı daha hızlı olarak azaltan bir üstel terim içerir (Azar vd., 2013). Bulanık en yüksek olabilirlik tahmini Öklid uzaklığı yerine üstel bir uzaklık fonksiyonu kullanır. Çünkü üstel uzaklık ölçüsü elips şeklindeki kümeler için daha uygundur (Ünler, 2006). Gath-Geva algoritması dar olan yerel bölgelerde bu üstel uzaklığa göre optimumu arar (Lange vd., 2006).

Gath-Geva algoritmasında temel düşünce, veri noktalarının p boyutlu normal dağılması varsayımına dayanır.

Bulanık maksimum benzerlik tahmincisi (fuzzy maximum likelihood estimate-*FMLE*) uzaklık fonksiyonu aşağıdaki gibidir.

$$D_{ij} = \frac{\sqrt{\det(F_i)}}{\alpha_i} exp\left(\frac{1}{2}(x_k - v_i)^T F_i^{-1}(x_k - v_i)\right) \tag{4.19}$$

F_i , bulanık kovaryans matrisi ise şöyle hesaplanmaktadır:

$$F_i = \frac{\sum_{j=1}^{n}(\mu_{ij})^m (x_j - v_i)(x_j - v_i)^T}{\sum_{j=1}^{n}(\mu_{ij})^m} \tag{4.20}$$

Fonksiyondaki α_i seçilen kümeden elde edilen ait önsel olasılıklardır.

$$\alpha_i = \frac{1}{n}\sum_{j=1}^{n}\mu_{ij} \tag{4.21}$$

Bulanık küme merkezleri ise aşağıdaki eşitliğe göre hesaplanır:

$$v_i = \frac{\sum_{j=1}^{n}(\mu_{ij})^m x_j}{\sum_{j=1}^{n}(\mu_{ij})^m} \tag{4.22}$$

Gath-Geva algoritması şu şekilde işlemektedir:

Adım 1: c küme sayısı, m bulanıklık katsayısı, i iterasyon sayısı, ε işlem sonlanma kriteri olmak üzere başlangıç değerleri belirlenir; μ_{ij} ‘lerden oluşan U üyelik matrisine rasgele değerler atanır.

Adım 2: (4.22) eşitliği ile her küme için küme merkezleri hesaplanır.

Adım 3: (4.20) eşitliği kullanılarak her küme için bulanık kovaryans matrisi hesaplanır.

Adım 4: (4.21) eşitliği ile önsel olasılıklar hesaplanır.

Adım 5: (4.19) eşitliği ile uzaklıklar hesaplanır.

Adım 6: $\left\| U_{yeni} - U_{eski} \right\| \leq \varepsilon$ ise iterasyon durdurulur. Aksi takdirde Adım 2’ye geri dönülür.

Gath-Geva algoritması olasılık teorisi temelleri üzerine kurulmuş bir algoritmadır. Gath-Geva algoritması, Bulanık c-ortalamalar ve Gustafson-Kessel algoritmalarının teşhis edebileceği tüm kümeleri başarıyla kümeler (Erilli, 2009). En büyük avantajı başlangıç prototipleri iyi seçilerek başlanması durumunda, eşit olmayan değişken özellikleri ve yoğunluklarında doğru bölünme sonuçları verebilmesidir (Lange vd., 2006). Bunun yanında yöntemin karmaşıklığı yüksektir ve kullanılan üstel terimli uzaklık ölçüsü nedeniyle yerel optimuma takılma olasılığı mevcuttur (Ünler, 2006; Azar vd., 2013).

4.3. Bulanık Kümeleme Geçerlilik İndeksleri

Kümeleme algoritmalarından kümeleme sonuçları elde edildikten sonra, verinin gerçek yapısını temsil edip etmediğini doğrulamak önemlidir. Küme geçerlilik indeksleri kümeleme algoritmasından elde edilen küme bölünmelerinin uygunluğunu değerlendirmek için kullanılır. Küme geçerliliği, elde edilen bölünmelerin doğruluğuna ve bu bölünmelerin doğruluğunun nasıl ölçülendirileceğine bakar. Kümeleme algoritması, kümeleri en iyi biçimde parametrelendirecek şekilde tasarlanmıştır (Azar vd., 2013).

Verideki en uygun küme sayısını belirlemek için iki ana yaklaşım vardır:

i. Yeteri kadar büyük küme sayısıyla başlayarak, ardı ardına bu sayısı azaltarak birbirine önceden tanımlanan kritere göre benzer olan kümeleri birleştirmektir. Bu yaklaşım *uyumlu kümeleri birleştirme* olarak adlandırılır.

ii. Kümelemeyi farklı c değerleri için yapmak ve belirlenen bölünmelerin iyiliğini değerlendirmek için geçerlilik indekslerini kullanmaktır. Bu iki farklı yolla yapılabilir:

- Birinci yol, tam bölünmeyi hesaplayan bir geçerlilik fonksiyonu tanımlamaktır. Küme sayısı için tahmini bir üst bant ($c_{\max}$) belirlenir ve algoritma her bir $c \in 2,3,\dots,c_{max}$ için çalıştırılır.
- İkinci yol ise; küme bölünmelerindeki kümeleri ayrı ayrı hesaplayan bir geçerlilik fonksiyonu tanımlamaktır. Burada yine bir $c_{\max}$ belirlenmeli ve kümeleme analizi $c_{\max}$ ile yürütülmelidir. Küme sonuçları belirlenen geçerlilik fonksiyonu esas alınarak birbirleriyle karşılaştırılır. Benzer kümeler tek bir kümede toplanır, çok kötü olan kümeler elenir, böylece küme sayısı azaltılır. Bu prosedür, kötü kümeler var oldukça devam ettirilebilir (Abonyi ve Feil, 2007).

4.3.1. Bölünme katsayısı (partition coefficient-PC)

En sıklıkla kullanılan ölçülerden biri olan bölünme katsayısı Bezdek (1981) tarafında tanımlanmıştır. Çakışan kümelerin miktarını ölçen bölünme katsayısı Bezdek tarafından şu şekilde tanımlanmıştır (Grekousis ve Thomas,2012; Balasko vd., 2005):

$$PC(c) = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij})^2 \tag{4.23}$$

Burada *c* küme sayısını gösterirken, indeks değerleri $[\mathbf{1/c}, \mathbf{1}]$ aralığındadır. *PC* değerinin *1*'e yaklaşıyor olması daha iyi bir kümeleme yapıldığına işaret eder (Grekousis ve Thomas, 2012; Wang ve Zhang, 2007).

PC'nin dezavantajı ise *c*'ye göre sürekli azalan bir fonksiyona sahip olması ve veriyle doğrudan doğruya bağlantılı olmamasıdır (Grekousis ve Thomas,2012; Abonji ve Feil, 2007; Azar vd., 2013).

4.3.2. Sınıflandırma entropisi (classification entropy-CE)

Yine Bezdek (1981) tarafından önerilen sınıflandırma entropisi, Bölünme katsayısına benzer olarak, küme bölünmelerindeki bulanıklığı ölçer. Aşağıdaki gibi tanımlanmıştır:

$$CE(c) = -\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}\ln(\mu_{ij}) \tag{4.24}$$

CE değerinin *0*'a yaklaşıyor olması daha iyi bir kümeleme yapıldığını gösterir. Dezavantajı ise; küme sayısı *c* artarken artan bir eğilime sahip olmasıdır. Sonuç olarak *PC* ve *CE*, küme sayısı *c* 'nin değerlerine göre monoton bir eğilim gösterirler (Grekousis ve Thomas,2012).

4.3.3. Bölünme indeksi (partition index-SC)

Bölünme indeksi kümelerdeki sıkılaşma ve ayrışmanın toplamlarının oranıdır.

$$SC(c) = \sum_{i=1}^{c} \frac{\sum_{j=1}^{n} (\mu_{ij})^m \|x_j - v_i\|^2}{n_i \sum_{k=1}^{c} \|v_k - v_i\|^2} \tag{4.25}$$

SC eşit sayıdaki kümelere sahip farklı bölünmeleri kıyaslamada kullanışlı bir indekstir. Düşük *SC* değeri daha iyi bir bölünme olduğunu gösterir (Abonyi ve Feil, 2007).

4.3.4. Ayrılma indeksi (seperation index-S)

Ayrılma indeksi, bölünme geçerliliğinde minimum uzaklık ayrılmalarını kullanır.

$$S(c) = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^2 \|x_j - v_i\|^2}{n min_{ik} \|v_k - v_i\|^2} \tag{4.26}$$

En küçük ayrılma indeksi, optimal bölünmeyi gösterir (Grekousis ve Thomas,2012).

4.3.5. Xie-Beni indeksi (Xie-Beni index-XB)

Xie- Beni indeksi üyelik dereceleri ile verideki geometrik yapıyı birleştirir. Xie-beni indeksi tüm sıkılaşmanın ortalamasına karşılık gelen ayrışmayı ölçer.

$$XB(c) = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m \|x_j - v_i\|^2}{n min_{ij} \|v_k - v_i\|^2} \tag{4.27}$$

Daha küçük Xie-Beni daha birleşik ve daha iyi ayrılmış bir küme anlamına gelir. Optimal kümeleme $c \in 2,3,\ldots,c_{max}$ olmak üzere c'ye göre Xie-Beni indeksinin minimize edilmesiyle elde edilir (Abonji ve Feil, 2007; Azar vd., 2013).

4.3.6. Dunn indeksi (Dunn's index-DI)

Dunn indeksi birleşik ve daha iyi ayrılmış kümeleri tanımlamak için önerilmiştir.

$$DI(c) = min_{i\in c}\left\{min_{j\in c, i\neq j}\left\{\frac{min_{x\in c_i, y\in c_j} d(x,y)}{max_{k\in c}\{max_{x,y\in c} d(x,y)\}}\right\}\right\} \tag{4.28}$$

Dunn indeksinin dezavantajı artan *c* ve *n* sayıları ile hesaplama için gereken sayısal gereksinimin genişlemesi ve hesaplamanın zorlaşmasıdır (Abonji ve Feil, 2007).

4.3.7. Alternatif Dunn İndeksi (Alternative Dunn's Index-ADI)

Alternatif Dunn indeksinin amacı, iki küme arasındaki uzaklık ölçüsü olarak üçgen eşitsizliği kullanarak, orijinal Dunn indeksindeki hesaplamayı daha basit hale getirmektir. Üçgen eşitsizliği;

$$d(x,y) \geq \left|d(y,v_j) - d(x,v_j)\right| \tag{4.29}$$

v_j , *j*. kümenin merkezi olmak üzere *ADI*,

$$ADI(c) = min_{i\in c}\left\{min_{j\in c, i\neq j}\left\{\frac{min_{x_i\in c_i, x_j\in c_j}\left|d(y,v_j) - d(x,v_j)\right|}{max_{k\in c}\{max_{x,y\in c} d(x,y)\}}\right\}\right\} \tag{4.30}$$

şeklinde hesaplanır (Abonji ve Feil, 2007).

Küme sayısı önceden bilinmiyorsa, geçerlilik indeksleri yararlı olacaktır. Bu durumda, *c*'nin farklı değerlerine göre kümeleme denemeleri yürütülür ve önceden tanımlanmış bu indekslere karşılık gelen grafikler çizilir. Optimum sonuç bu indekslerin ve grafiklerinin karşılaştırmalı olarak incelenmesiyle bulunur (Azar vd., 2013).

PC ve *CE* indekslerinin her ikisi de monoton eğilimlidir. Optimum küme sayısı *PC* için artan eğilim ve *CE i*çin azalan eğilim, daha basitçe anlatırsak, bu iki indeksin grafiklerindeki en keskin değişimdir. *SC, S* ve *XB* indekslerinde optimum küme sayısı bu indeksler minimum olduğundadır. Bunun yanında, *DI* ve *ADI* indeksleri çakışan kümelerde güvenilir sonuçlar vermemektedir. Unutulmamalıdır ki; hiçbir geçerlilik indeksi tek başına optimum küme sayısını belirlemek için yeterli ve güvenilir değildir. Her zaman doğru sonucu veren bir yöntem yoktur (Azar vd., 2013; Abonji ve Feil, 2007).

5. UYGULAMA

5. 1. OECD Ülkelerinin Gelişmişlik Bakımından Kümelendirilmesi

Uygulamada Ekononomik İşbirliği ve Kalkınma Örgütü (OECD- The Organization for Economic Co-operation and Development) ülkelerinin sosyo-ekonomik yapılarını gösteren 30 değişkene ait veriler kullanılmıştır. Ülkelerin ekonomik performansını gösteren makro-ekonomik ve sosyo-kültürel göstergeler ile örgütlenmede yer alan bu ülkelerin gelişmişlik bakımından hangileriyle benzeştiğinin, hangileriyle farklılaşma gösterdiğinin ortaya konulması hedeflenmiştir.

OECD, 14 Aralık 1960 yılında imzalanan Paris Sözleşmesi'ne dayanılarak 1961 yılında Avrupa Ekonomik İşbirliği Örgütü (Organization for European Economic Co-operation)' nün mirasçısı olarak kurulan ve çoğunluğunu Avrupa Birliği ülkelerinin oluşturduğu bir kuruluştur. Örgütün amacı anlaşmanın birinci maddesinde;

i. Mali istikrarı koruyarak, en yüksek sürdürülebilir ekonomik büyümeyi ve istihdamı sağlayacak, üye ülkelerde hayat standardını yükseltecek ve böylece dünya ekonomisinin gelişmesine,
ii. Üye olan veya olmayan ülkelerde sağlıklı ekonomik kalkınmaya,
iii. Uluslararası yükümlülüklere uygun olarak çok taraflı dünya ticaretinin büyümesine ayırım yapmadan katkıda bulunacak siyasalar geliştirmek olarak belirtilmektedir (MFA, 2015).

OECD'nin 20 kurucu üyesi bulunmaktadır:

ABD, Avusturya, Kanada, Fransa, Hollanda, Lüksemburg, Almanya, İtalya, İngiltere, Belçika, Danimarka, İrlanda, Yunanistan, İsviçre, İsveç, İspanya, İzlanda, Norveç, Portekiz ve Türkiye.Bunların yanında ilerleyen yıllarda kuruluşa bazı ülkeler “üye” olarak katılmıştır: Japonya, Finlandiya, Avustralya, Güney Kore, Meksika ve Yeni Zelanda.

Sovyetler Birliği’nin dağılmasından sonra kuruluşa dâhil olan ülkeler ise; Polonya, Macaristan, Çek Cumhuriyeti ile Slovakya’ dır. Son olarak 2010 yılında Estonya, İsrail, Slovenya ve Şili’nin de katılımıyla OECD’ ye üye ülke sayısı 34 ‘tür. (OECD, 2015; MFA, 2015). OECD’ye üye ülkelere ait çizelge Çizelge 5.1’ de verilmiştir.

Çizelge 5. 1. OECD Üyesi Ülkeler

Kurucu Üyeler	Sonradan Dâhil Olan Ülkeler
Almanya	Avustralya
ABD	Çek Cumhuriyeti
Avusturya	Estonya
Belçika	Finlandiya
Büyük Britanya	İsrail
Danimarka	Güney Kore
Fransa	Japonya
Hollanda	Macaristan
İrlanda	Meksika
İtalya	Polonya
İspanya	Slovakya
İsveç	Slovenya
İsviçre	Şili
İzlanda	Yeni Zelanda
Kanada	
Lüksemburg	
Norveç	
Portekiz	
Türkiye	
Yunanistan	

5. 2. Uygulamada Kullanılan Sosyoekonomik Değişkenler

Çizelge 5. 2. Uygulamada kullanılan sosyoekonomik değişkenler

Değişken No	Değişken Adı
X_1	Nüfus artış hızı (yüzde)
X_2	GSYİH’ dan sağlığa ayrılan pay (yüzde)
X_3	Bin kişiye düşen doktor sayısı
X_4	Kişi başına sağlık harcaması (ABD Doları)
X_5	Toplam doğurganlık hızı
X_6	Bebek ölüm hızı
X_7	GSYİH’ nın yüzdesi olarak ticaret dengesi
X_8	Toplam istihdamın yüzdesi olarak part- time istihdam
X_9	Toplam istihdamın yüzdesi olarak kendi işinde istihdam
X_{10}	İşsizlik oranı
X_{11}	Toplam istihdamın yüzdesi olarak tarımda istihdam
X_{12}	Toplam istihdamın yüzdesi olarak endüstride istihdam
X_{13}	Toplam istihdamın yüzdesi olarak hizmetlerde istihdam
X_{14}	GSYİH’ nın yüzdesi olarak vergi oranı
X_{15}	Doğuşta yaşam beklentisi
X_{16}	Yetişkin okuryazar oranı
X_{17}	Kentsel nüfus oranı
X_{18}	GSMH’ nın yüzdesi olarak kamunun eğitim harcaması
X_{19}	Bin kişiye düşen cep telefonu aboneliği
X_{20}	Bin kişiye düşen internet aboneliği
X_{21}	GSMH’ nın yüzdesi olarak ar-ge harcaması
X_{22}	Satın alma gücü paritesine göre kişi başına düşen milli gelir
X_{23}	Kişi başına GSYİH artış hızı
X_{24}	TÜFE artış hızı
X_{25}	GSYİH’nın yüzdesi olarak mal ve hizmet ithalatı
X_{26}	GSYİH’nın yüzdesi olarak mal ve hizmet ihracatı
X_{27}	Mamul ürün ihracatının yüzdesi olarak ileri teknoloji ihracatı
X_{28}	GSYİH’ nın yüzdesi olarak net doğrudan yabancı sermaye girişi
X_{29}	GSYİH’ nın yüzdesi olarak askeri harcamalar
X_{30}	15+ yaş kadınların ekonomiye katılım oranı

Uygulamada kullanılan 34 OECD ülkesine ait 30 sosyoekonomik değişken Çizelge 5.2' de verilmiştir.

Uygulama kullanılan veri seti; 30 değişkene ait 2010 ve sonrası güncel verilerden; OECD, EUROSTAT, CIA, Worldbank, IMF, UN'nin internet sitelerinde yer alan veri bankaları aracılığıyla ve ILO 'ya ait KILM (Key Indicators of Labour Market) yazılımından elde edilmiştir.

OECD üyesi 34 ülkenin sosyoekonomik gelişmişlik göstergelerine ait 30 değişkenden oluşan veri setine, ülkelerin gelişmişlik bakımından oluşturduğu kümelenmelerin tespit edilmesi amacıyla geleneksel bulanık kümeleme yöntemlerinden bulanık c-ortalamalar kümeleme yöntemi uygulanmıştır. Ayrıca aynı veri setine hiyerarşik olmayan kümeleme yöntemlerinden, aykırı değerlere karşı duyarlı olmaması bakımından bulanık c- ortalamalar yöntemiyle karşılaştırmada tercih edilen k- medoid kümeleme yöntemi uygulanmış ve bu iki kümeleme yöntemi sonucu elde edilen kümelenmeler karşılaştırılmıştır.

Uygulama için öncelikle MATLAB paket programı kullanılmıştır. Programa eklenen "Fuzzy Clustering Toolbox- FUZZCLUST" kullanılarak bulanık c-ortalamalar kümeleme algoritması ve yine programa eklenen "Som Toolbox- SOM" aracı ile verinin kümelenmesini gösteren grafik algoritması yazılmıştır. Daha sonra aynı veri setine, R paket programına eklenen "e1071" paketi kullanılarak bulanık c-ortalamalar kümeleme yöntemi uygulanmış ve kümeleme sonuçları elde edilmiştir. K-medoid kümeleme uygulaması için ise NCSS 2007 paket programı kullanılmıştır.

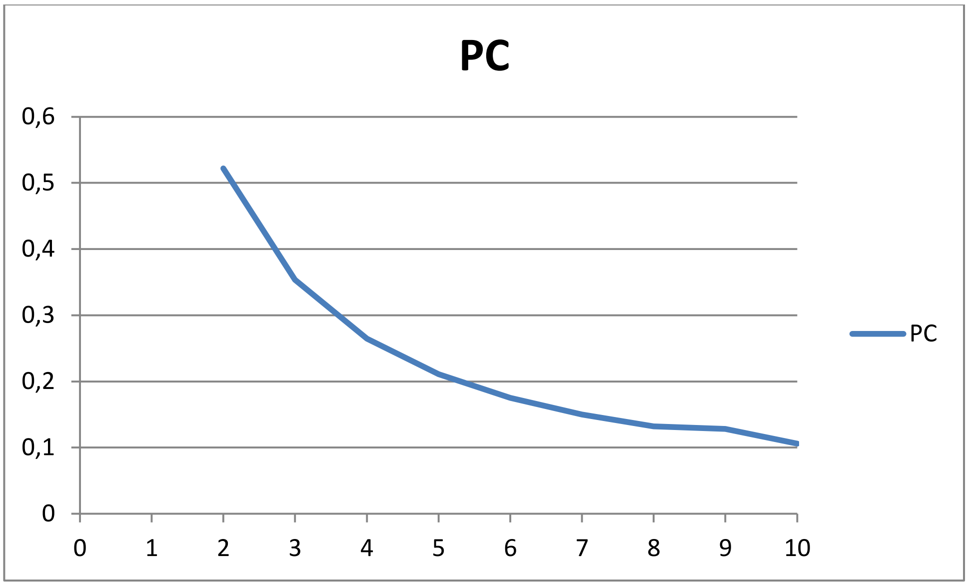
6. BULGULAR VE TARTIŞMA

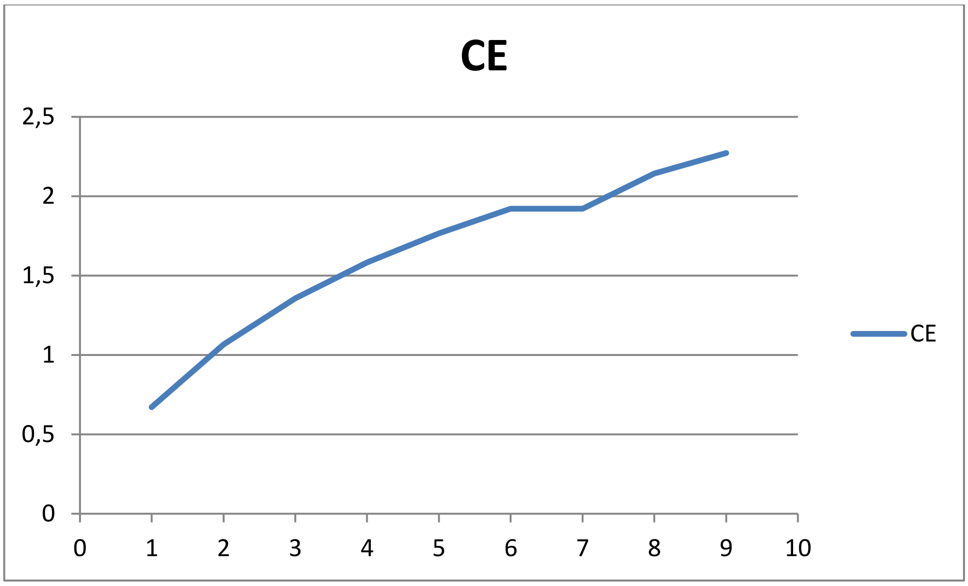
En uygun küme sayısının belirlenmesi için, öncelikle Şahin ve Hamarat (2002)‘ ın çalışmasından yola çıkılarak OECD ülkelerini sosyoekonomik gelişmişlik açısından çok gelişmiş ve daha az gelişmiş olarak iki ayrı gruba ayırmak hedeflenmiştir. Ek olarak Bölüm 4.3.1 ve Bölüm 4.3.2’ de bahsedilen ve literatürde bulanık c-ortalamalar yöntemi için en çok kullanılan geçerlilik indekslerinden bölünme katsayısı (*PC*) ve sınıflandırma entropisi (*CE*) değerlerine (Çizelge 6.1) bakılarak da en uygun küme sayısının “iki” olacağına karar verilmiştir. Bunun için küme sayısı “*2*” ile “*10*” arasında değiştirilerek oluşan *PC* ve *CE* değerleri kıyaslanmıştır.

Literatürde *PC* değerinin *1*’e yakın olması ve *CE* değerinin de *0*’a yakın olması istenmektedir. Uygulamalarda ise *PC* ve *CE* değerinin birbirine en yakın olduğu küme sayısı en uygun olarak kabul edilmektedir. Ayrıca grafik incelemelerinde ise *PC* değeri için, iki *PC* değeri arasındaki farkın en büyük olduğu ilk yerel maksimum ve *CE* değeri için ise iki *CE* değeri arasındaki farkın en büyük olduğu ilk yerel minimum da tercih edilebilir. Çalışmada *PC* ve *CE* değerinin birbirlerine en yakın olduğu değer olan ve ayrıca *PC* değerinin *1*’e en yakın ve *CE* değerinin de *0*’a en yakın olduğu “iki” en uygun küme sayısı olarak tespit edilmiştir.

Çizelge 6. 1. *PC* ve *CE* geçerlilik indeks değerleri

Küme Sayısı	**PC**	**CE**
2	0,521728955	0,671179948
3	0,353874758	1,06678069
4	0,264661331	1,355990505
5	0,210926217	1,58146157
6	0,17535594	1,765475461
7	0,150268103	1,920367579
8	0,132344159	1,920367579
9	0,128284391	2,143401867
10	0,106139378	2,270822564

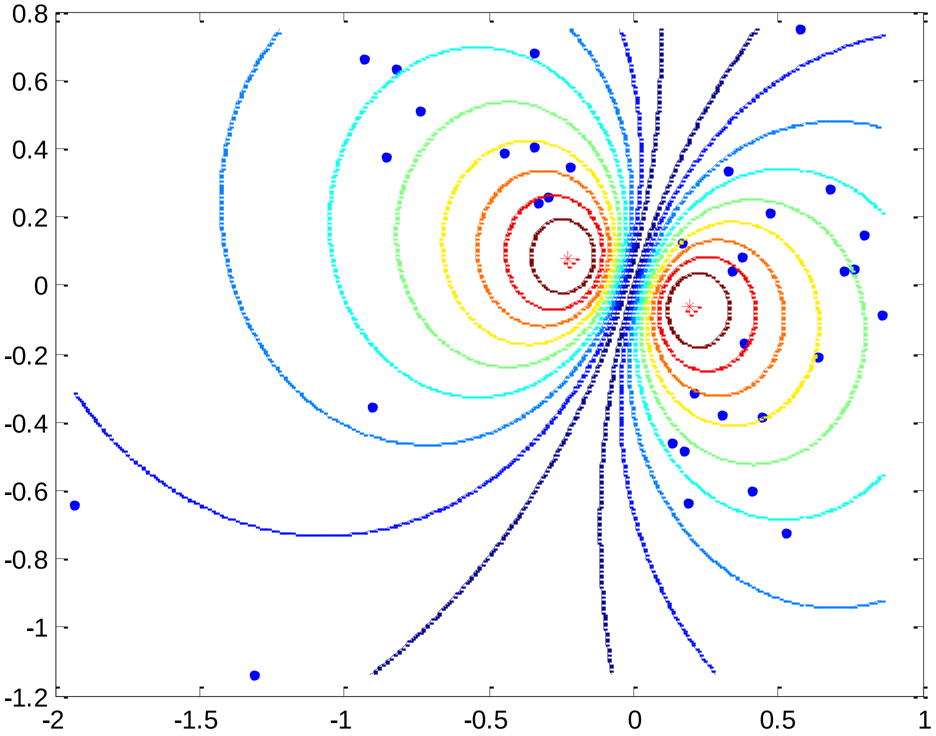




Şekil 6. 1. *PC* ve *CE* geçerlilik indekslerine ait grafikler

PC ve *CE* geçerlilik indeks değerleri (Bkz. Çizelge 6.1) ve *PC* ve *CE* geçerlilik indeks değerlerine ait grafikler (Şekil 6.1) incelendiğinde; küme sayısı “2” iken *PC* ve *CE* değerleri birbirlerine en yakın değerleri almakta ve *PC*’nin, *1*’e en yakın değeri ve *CE*‘nin ise *0*’a en yakın değeri küme sayısı “2” olduğunda aldığı açıkça görülmektedir.

Küme sayısı “2” alınarak MATLAB paket programında oluşturulan algoritma yardımıyla yapılan Bulanık c- Ortalamalar kümeleme sonucu elde edilen küme dağılım grafiği Şekil 6. 2’deki gibidir.



Şekil 6. 2. MATLAB’ de Bulanık-c Ortalamalar sonucu elde edilen kümeleme dağılım grafiği

MATLAB programı ile elde edilen, OECD ülkelerinin üyelik değerleri ve üyelik değerlerine bakılarak hangi kümede yer aldıkları ise Çizelge 6.2 ‘de verilmiştir. Birimler hangi kümeye daha büyük küme üyelik değerine sahipse, o kümeye ait olacak şekilde kümelere atama yapılmaktadır.

Çizelge 6. 2. MATLAB ile elde edilen OECD ülkelerine ait üyelik değerleri

	Kümeler	
ÜLKELER	**1**	**2**
Almanya	0,63022805	0,36977195
ABD	0,57530455	0,42469545
Avustralya	0,64019344	0,35980656
Avusturya	0,58021914	0,41978086
Belçika	0,60398777	0,39601223
Büyük Britanya	0,64092584	0,35907416
Çek Cumhuriyeti	0,37042111	0,62957889
Danimarka	0,61495396	0,38504604
Estonya	0,42661988	0,57338012
Finlandiya	0,58392679	0,41607321
Fransa	0,6183706	0,3816294
Hollanda	0,60964478	0,39035522
İrlanda	0,51693541	0,48306459
İspanya	0,4239081	0,5760919
İsrail	0,54187016	0,45812984
İsveç	0,67866542	0,32133458
İsviçre	0,61559487	0,38440513
İtalya	0,41608079	0,58391921
İzlanda	0,65791915	0,34208085
Japonya	0,59955588	0,40044412
Kanada	0,59704286	0,40295714
Güney Kore	0,53128551	0,46871449
Lüksemburg	0,51817246	0,48182754
Macaristan	0,39497769	0,60502231
Meksika	0,44808736	0,55191264
Norveç	0,59604197	0,40395803
Polonya	0,33942144	0,66057856
Portekiz	0,39183187	0,60816813
Slovakya	0,3657887	0,6342113
Slovenya	0,39197749	0,60802251
Şili	0,42074845	0,57925155
Türkiye	0,43655216	0,56344784
Yeni Zelanda	0,59200431	0,40799569
Yunanistan	0,41685575	0,58314425

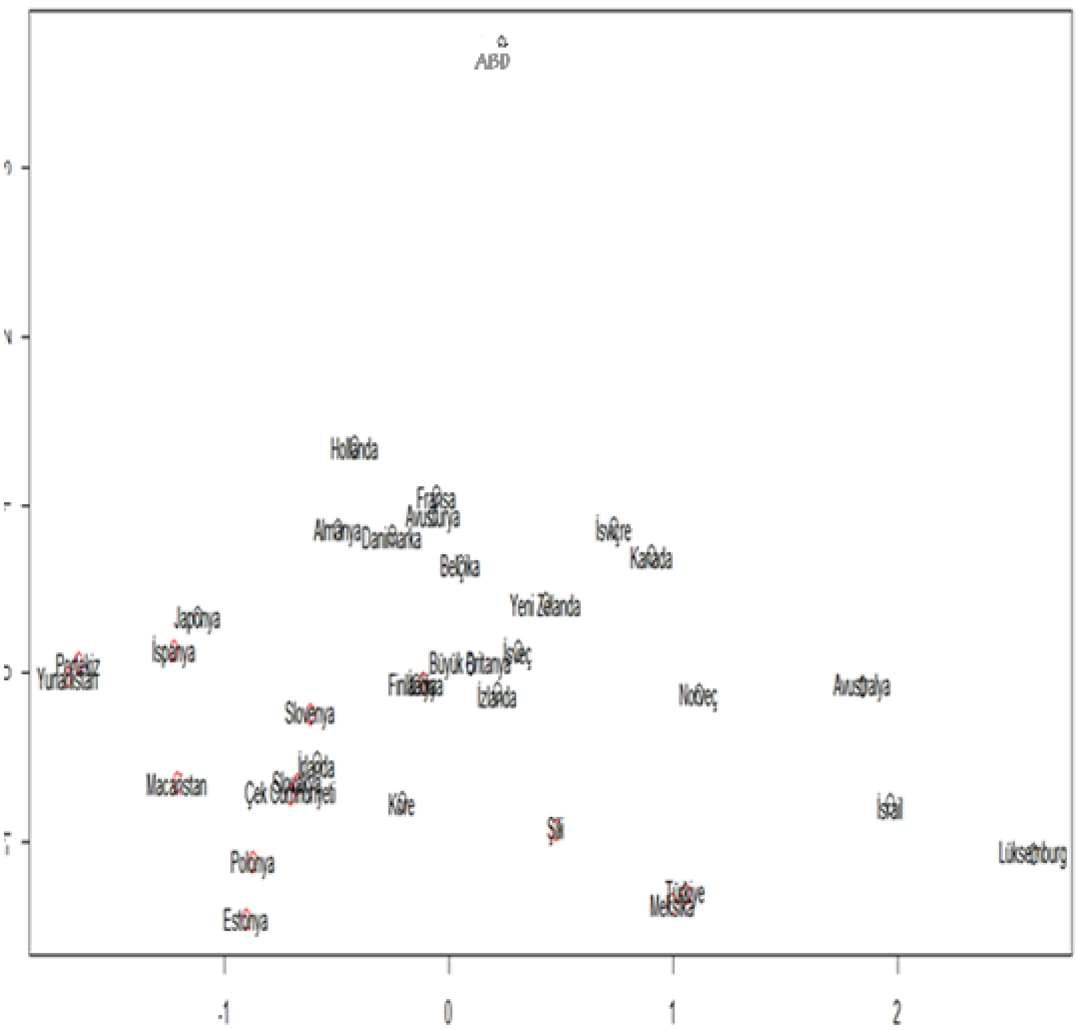
MATLAB programı kullanılarak, Bulanık c-Ortalamalar Kümeleme Yöntemi ile OECD ülkeleri (34 ülke) küme üyelik değerlerine bakılarak çok gelişmiş ve daha az gelişmiş olmak üzere iki ayrı grupta kümelendirilmiştir. Bu kümeleme sonucunda 34 ülkeden 21'i sosyoekonomik bakımdan çok gelişmiş ülkeler grubunda yer alırken; Türkiye'nin de içinde bulunduğu diğer grupta yer alan 13 ülke ise OECD üyesi ülkeler içinde daha az gelişmiş ülkeler olarak gruplanmıştır (Çizelge 6.3).

Çizelge 6. 3. Bulanık c-Ortalamalar Kümeleme sonucu OECD ülkeleri ve yer aldığı kümeler

Gelişmiş Ülkeler	**Daha Az Gelişmiş Ülkeler**
Almanya	Çek Cumhuriyeti
ABD	Estonya
Avustralya	İspanya
Avusturya	İtalya
Belçika	Macaristan
Büyük Britanya	Meksika
Danimarka	Polonya
Finlandiya	Portekiz
Fransa	Slovakya
Hollanda	Slovenya
İrlanda	Şili
İsrail	Türkiye
İsveç	Yunanistan
İsviçre	
İzlanda	
Japonya	
Kanada	
Güney Kore	
Lüksemburg	
Norveç	
Yeni Zelanda	

Gelişmiş ülkeler incelendiğinde kişi başına düşen sağlık harcaması, işsizlik oranı, kişi başına düşen milli gelir, TÜFE artış hızı, ileri teknoloji ihracatı gibi değişkenler için aldıkları değerlerin daha az gelişmiş ülkelere kıyasla çok daha iyi durumda olduğu gözlemlenmektedir. Daha az gelişmiş ülkelerin sağlık, milli gelir, işsizlik, ihracat gibi gelişmişlik göstergesi bu alanlarda gelişmiş ülkeler kümesine göre çok daha geride kalmaktadır.

Benzer biçimde yine aynı veri seti kullanılarak R paket programıyla OECD ülkesi 34 ülke sosyoekonomik gelişmişlik bakımından iki ayrı küme olacak şekilde bulanık c-ortalamalar kümeleme analizi ile incelenmiştir. Analiz sonu elde edilen kümeleme grafiği Şekil 6. 3'teki gibidir.



Şekil 6. 3. R' da Bulanık c-Ortalamalar Kümeleme sonucu elde edilen kümeleme grafiği

Çizelge 6. 4. R ile elde edilen OECD ülkelerine ait üyelik değerleri

	Kümeler	
ÜLKELER	**1**	**2**
Almanya	0.484432113018321	0.515567886981679
ABD	0.490400210269798	0.509599789730202
Avustralya	0.47912369061752	0.52087630938248
Avusturya	0.490366408211764	0.509633591788236
Belçika	0.487281723763	0.512718276237
Büyük Britanya	0.479181832666835	0.520818167333165
Çek Cumhuriyeti	0.519945885408802	0.480054114591198
Danimarka	0.484113048209745	0.515886951790255
Estonya	0.510794940195574	0.489205059804426
Finlandiya	0.489203242816809	0.510796757183191
Fransa	0.48377534915293	0.51622465084707
Hollanda	0.486431456506529	0.513568543493471
İrlanda	0.498357902564714	0.501642097435286
İspanya	0.510474098602961	0.489525901397039
İsrail	0.494825437083196	0.505174562916804
İsveç	0.475185119997749	0.524814880002251
İsviçre	0.484050372549085	0.515949627450915
İtalya	0.512275084564241	0.487724915435759
İzlanda	0.480141307891555	0.519858692108445
Japonya	0.488864026234955	0.511135973765045
Kanada	0.486285508776757	0.513714491223243
Güney Kore	0.497404097982757	0.502595902017243
Lüksemburg	0.499094815002843	0.500905184997157
Macaristan	0.51656023818167	0.48343976181833
Meksika	0.506109631285632	0.493890368714368
Norveç	0.48653256559042	0.51346743440958
Polonya	0.522133250050947	0.477866749949053
Portekiz	0.514024599836508	0.485975400163492
Slovakya	0.519168528328924	0.480831471671076
Slovenya	0.516116924497276	0.483883075502724
Şili	0.509700871672713	0.490299128327287
Türkiye	0.507594899818315	0.492405100181685
Yeni Zelanda	0.483625597006161	0.516374402993839
Yunanistan	0.510090802118177	0.489909197881823

R paket programı kullanılarak elde edilen OECD ülkeleri küme üyelik değerleri ve bu değerler kullanılarak yer aldıkları kümeler ise Çizelge 6.4' teki gibidir.

R sonucu elde edilen küme üyelikleri, MATLAB programından farklı olarak birbirlerine daha yakın sonuçlar (daha bulanık sonuçlar) vermekle birlikte nihai olarak elde edilen kümelenmede ülkelerin yer aldıkları gruplanmalar (Çizelge 6.5) değişmemiştir.

Çizelge 6. 5. R ile yapılan bulanık c-ortalamalar analizi sonucu OECD ülkelerinin yer aldığı kümeler

Gelişmiş Ülkeler	**Daha Az Gelişmiş Ülkeler**
Almanya	Çek Cumhuriyeti
ABD	Estonya
Avustralya	İspanya
Avusturya	İtalya
Belçika	Macaristan
Büyük Britanya	Meksika
Danimarka	Polonya
Finlandiya	Portekiz
Fransa	Slovakya
Hollanda	Slovenya
İrlanda	Şili
İsrail	Türkiye
İsveç	Yunanistan
İsviçre	
İzlanda	
Japonya	
Kanada	
Güney Kore	
Lüksemburg	
Norveç	
Yeni Zelanda	

Bununla birlikte R programıyla elde edilen kümeleme grafiğinde ülkeler etiketlenerek hangi ülkenin nerede yer aldığını görmek kolaylaşmıştır.

Burada (Bkz. Şekil 6.3) yardımıyla, özelikle Portekiz ile Yunanistan, Slovakya ile Çek Cumhuriyeti, Fransa ile Avusturya ve son olarak Türkiye ile Meksika'nın büyük bir benzerlik gösterdiği; bunun yanında ABD'nin konumlanmasının diğer OECD ülkelerine göre farklı bir yerde bulunduğunu söyleyebiliriz.

Son olarak MATLAB ve R ile elde edilen kümelerin doğru kümelenip kümelenmediğini belirlemek amacıyla IBM SPSS Statistics 22 paket programı yardımıyla elde edilen kümelerle birlikte veri setine Diskriminant Analizi uygulanmıştır.

Diskriminant analizi ile elde edilen kovaryans matrislerinin eşitliği Box's M testi (Çizelge 6.6) ile test edilmiştir. Box's M testi sonucuna göre grup kovaryans matrisleri eşittir. Doğrusal diskriminant analizi uygulanabilir.

Çizelge 6. 6. Box's M testi sonuçları

Box's M		1,399
F	Yaklaşık	1,353
	df1	1
	df2	2590,920
	Anlamlılık	,245

Wilks' Lambda (Çizelge 6.7) değerine bakıldığında diskriminant fonksiyonunun ayırma gücünün önemli düzeyde yüksek olduğu görülmektedir ($p<0.01$).

Çizelge 6. 7. Wilks' Lambda sonuçları

Wilks' Lambda

Fonksiyon Testi	Wilks' Lambda	Chi-square	df	Anlamlılık
1	,022	72,873	26	,000

Diskriminant analizi ile elde edilen çapraz doğrulama sonucunda (Çizelge 6.8) ülkeler gruplara %100 oranda doğru atanmıştır.

Çizelge 6. 8. Diskriminant analizi çapraz doğrulama sonucu

Sınıflandırma Sonuçları

		Grup	Tahmini Grup Üyeliği 1,00	Tahmini Grup Üyeliği 2,00	Toplam
Orijinal	Gruplanan	1,00	21	0	21
		2,00	0	13	13
	%	1,00	100,0	,0	100,0
		2,00	,0	100,0	100,0

Bulanık kümeleme ile klasik kümeleme sonuçlarını karşılaştırmak amacıyla klasik kümeleme yöntemlerinden olan k- medoid kümeleme yöntemi de kullanılmıştır. k-medoid uygulaması için NCSS' de veri girişi yapıldıktan sonra, analiz kısmında yer alan "Clustering" aracından "Medoid Partitioning" seçilmiş, en uygun küme sayısının belirlenmesi amacıyla küme sayısı minimum 2 ve maksimum 10 olarak giriş yapılmıştır. Kümelere ilişkin elde edilen ortalama gölge istatistiği (Çizelge 6.9) değerlerine bakılarak, ortalama gölge istatistiği (SC) değerinin en büyüklendiği küme sayısı k=2 en uygun olarak alınmıştır.

Çizelge 6. 9. Kümelere ilişkin ortalama gölge istatistiği

Küme Sayısı	Ortalama Uzaklık	Düzeltilmiş Ortalama Uzaklık	Ortalama Gölge İstatistiği
2	329,129675	19,360569	0,187247
3	203,839280	17,985819	0,105084
4	142,125956	16,720701	0,092756
5	107,048073	15,742364	0,074856
6	83,752704	14,779889	0,096938
7	68,342915	14,070600	0,085993
8	56,062439	13,191162	0,086314
9	47,153404	12,481784	0,060900
10	41,742367	12,277167	0,035318

Küme sayısı k=2 olarak alındığında, OECD ülkeleri kümelere kesin olarak atanmıştır. Ülkelerin atanması sonucu elde edilen kümeler Çizelge 6.10’ da verilmiştir:

Çizelge 6. 10. k-medoid kümeleme sonucu OECD ülkelerinin yer aldığı kümeler

Birinci Küme	**İkinci Küme**
Almanya	Çek Cumhuriyeti
ABD	Estonya
Avustralya	İrlanda
Avusturya	İspanya
Belçika	İtalya
Büyük Britanya	Güney Kore
Danimarka	Lüksemburg
Finlandiya	Macaristan
Fransa	Meksika
Hollanda	Polonya
İsrail	Portekiz
İsveç	Slovakya
İsviçre	Slovenya
İzlanda	Şili
Japonya	Türkiye
Kanada	Yunanistan
Norveç	
Yeni Zelanda	

k- medoid kümeleme ile OECD ülkeleri gelişmişlik bakımından “2” kümeye ayrılmıştır. Birinci kümeye bakıldığında OECD üye ülkeleri içinde gelişmiş ülkelerin var olduğu, ikinci kümede ise daha az gelişmiş üyelerin bulunduğu gözlemlenmektedir. Bu açıdan değerlendirildiğinde; OECD ülkeleri birinci kümede yer alan 18 gelişmiş ülke ve ikinci kümede yer alan 16 daha az gelişmiş ülke olmak üzere “2” kümeye ayrılmıştır.

Elde edilen kümelerin doğru kümelenip kümelenmediğini belirlemek amacıyla veri setine Diskriminant Analizi uygulanmıştır. Diskriminant analizi ile elde edilen kovaryans matrislerinin eşitliği Box's M testi (Çizelge 6.11) ile test edilmiştir. Box's M testi sonucuna göre grup kovaryans matrisleri eşittir. Doğrusal diskriminant analizi uygulanabilir.

Çizelge 6. 11. Box's M testi

Box's M		,147
F	Yaklaşık	,143
	df1	1
	df2	3040,125
	Anlamlılık	,705

Wilks' Lambda (Çizelge 6.12) değerine bakıldığında diskriminant fonksiyonunun ayırma gücünün önemli düzeyde yüksek olduğu görülmektedir ($p<0.01$).

Çizelge 6. 12. Wilks' Lambda sonuçları

Wilks' Lambda

Fonksiyon Testi	Wilks' Lambda	Chi-square	df	Anlamlılık
1	,019	71,677	28	,000

Çizelge 6. 13. Diskriminant analizi çapraz doğrulama sonucu

Sınıflandırma Sonuçları

		Grup	1,00	2,00	
Orijinal	Gruplanan	1,00	16	0	16
		2,00	0	18	18
	%	1,00	100,0	,0	100,0
		2,00	,0	100,0	100,0

Diskriminant analizi ile elde edilen çapraz doğrulama sonucunda (Çizelge 6.13) ise ülkeler gruplara %100 oranda doğru atanmıştır.

Bulanık c-ortalamalar sonucunda elde edilen kümelerle, k- medoid kümeleme sonucu elde edilen kümeler karşılaştırıldığında İrlanda, Kore ve Lüksemburg'un k- medoid kümeleme sonucunda daha az gelişmiş ülkeler grubunda yer aldığı gözlemlenmektedir. Bu noktada, bulanık c-ortalamalar kümeleme yöntemiyle elde edilen küme üyeliklerine bakacak olursak; İrlanda, Güney Kore ve Lüksemburg ülkelerinin kümelere olan üyelik derecelerinin birbirlerine çok yakın olduğu gözlemlenmektedir (Çizelge 6.14) .

Bilindiği üzere k-medoid kümeleme yöntemi kesin küme atamaları yapmaktadır. Öyle ki; burada bir ülkenin atandığı kümeye üyeliği "1" değerini alırken, atanmadığı küme için üyelik değeri "0" olmaktadır. Oysaki; bulanık kümeleme yöntemi ile elde edilen ülkelere dair üyelik değerleri yardımıyla, bu ülkelerin aslında gelişmiş ülkeler kümesine olan üyeliğinin daha fazla olduğu gözlemlenmektedir. Bu yüzden bulanık c-ortalamalar kümeleme yöntemi üyelik değerleri yardımıyla, k-medoid kümeleme yöntemine kıyasla daha kullanışlı ve işe yarar sonuçlar sunmaktadır.

Çizelge 6. 14. İrlanda, Kore ve Lüksemburg üyelik değerleri

	1	**2**
İrlanda	0,51693541	0,48306459
Güney Kore	0,53128551	0,46871449
Lüksemburg	0,51817246	0,48182754

7. SONUÇ VE ÖNERİLER

Çoğunluğu AB üyesi de olan gelişmiş Batı ülkelerinin içinde bulunduğu OECD üyesi 34 ülke, yapılan çalışma ile sosyoekonomik verilerden yola çıkılarak "çok gelişmiş" ve "daha az gelişmiş" olmak üzere iki farklı gruba ayrılmıştır. Birinci kümede, gelişmişlik göstergeleri bakımından AB ve OECD'nin zengin olarak kabul edilen ülkeleri yer alırken; ikinci grupta ise kurucu üyeler dışında olan ve OECD'ye sonradan dâhil olmuş ülkeler yer almaktadır. Türkiye'nin de içinde bulunduğu bu ikinci kümeye dâhil ülkeler AB ve OECD'nin daha az gelişmiş olan üyeleridir. Türkiye de dâhil olma üzere, ikinci kümede yer alan ülkelerin OECD üyeliklerinin bu noktada gelişmişlikten ziyade politik anlam ifade ettiği düşünülebilir.

Bunun yanında, Şahin ve Hamarat (2002)'ın "G-10, Avrupa Birliği ve OECD Ülkelerinin Benzerliklerinin Fuzzy Kümeleme Analizi ile Belirlenmesi" adlı çalışmalarında belirledikleri kümeleme sonucunda OECD'nin "daha az gelişmiş" ülkeleri grubunda yer alan Güney Kore'nin, çalışmamızda "gelişmiş ülkeler" grubunda yer aldığı gözlemlenmiştir. Ülkelerin sosyoekonomik gelişmişliklerin gösteren değişkenler kullanılarak oluşturulan veri seti incelendiğinde, Güney Kore' nin önceki yıllara göre sağlık, istihdam, ar-ge, kişi başına düşen milli gelir ve ithalat/ ihracat istatistikleri bakımından değerlerini arttırarak gelişmiş ülkeler seviyesine yükseldiği görülmüştür. Bunun başlıca nedenleri arasında, Güney Kore'nin bilgi ve yeni teknolojiler üreten inovasyon faaliyetlerindeki artış ve buna ek olarak, 1 Temmuz 2011 tarihinde AB ve Güney Kore arasında yürürlüğe giren "Serbest Ticaret Anlaşması- STA" gösterilebilir (Akses, 2010; Oğuztürk, 2011; Doğan ve Uzun, 2014).

AB üyeliği açısından Türkiye'nin durumu ele alındığında, Türkiye'yle birlikte ikinci kümede yer alan çoğu ülkenin (Çek Cumhuriyeti, Estonya, Macaristan, Polonya, Slovakya ve Slovenya) 2004 yılında AB üyeliğine giren ülkeler olduğu gözlemlenmektedir. Türkiye ise AB'ye aday bir ülke olarak 2004 yılından bu yana müzakere sürecini devam ettirmektedir. Bilindiği üzere, AB'ye tam üye olabilmek için yürütülen müzakereler doğrultusunda, aday ülkenin yapması gerekenler ve yerine getirmesi gereken koşullar belirlenmektedir. Bu süreçte AB'ye üyelik isteyen ülke, AB mevzuatına uyum sağlamalı, AB ortak politikasına uymalı ve yapılan anlaşmanın gereklerini yerine getirmelidir. Türkiye'ye adaylık statüsünün tanınmasıyla birlikte Türkiye'deki gelişmeler düzenli ilerleme raporları ile takip edilmeye başlanmıştır (Cavlak ve Bostanoğlu, 2015).

OECD ülkelerinin sosyoekonomik gelişmişliklerini gösteren değişkenlere ait veriler incelendiğinde; Türkiye' nin kişi başına düşen sağlık harcaması bakımından diğer ülkelere kıyasla oldukça düşük bir seviyede olduğu, doğurganlık hızı yüksek iken aynı zamanda da bebek ölüm hızının da fazla olduğu, kamunun eğitim harcamalarının diğer ülkelere göre daha az olduğu, ar- ge harcamalarında Türkiye' nin diğer ülkelerden geride kaldığı, TÜFE artış hızının fazla olduğu ve kadınların ekonomiye katkısı bakımından oldukça geride kaldığı açıkça görülmektedir. Bu veriler göz önüne alındığında Türkiye'nin neden OECD ülkeleri arasında "daha az gelişmiş ülkeler" kümesinde yer aldığı daha net anlaşılmaktadır.

Türkiye-Avrupa Birliği (AB) ilişkileri coğrafi, kültürel, ekonomik ve siyasi derinliğe sahip ilişkilerdir. Bu ilişkiler özünde stratejik ortaklığa dayanan, çok-boyutlu, kültürel, siyasi ve ekonomik maliyeti, katkısı ve getirisi yüksek, sistem-dönüştürücü niteliktedir. Kültürel düzeyde bakılırsa, Türkiye-AB ilişkileri Avrupa kimliğinin çok kültürlü hale gelmesinin yanında, Türk modernitesinin demokratik ve çoğulcu bir nitelik kazanmasını da talep etmektedir. Siyasi düzeyde, Türkiye'nin AB'nin tam üyesi olmasını içeren bütünleşme sürecinde Türkiye'de devletin demokratik, etkin, şeffaf ve sorumlu olmasını talep eden ilişkilerdir. Ekonomik düzeydeyse, bu ilişkiler AB ülkeleri Türk ekonomisinin en önemli hareket alanını oluşturmaktadırlar. AB'nin Türkiye'nin ekonomik olarak büyümesine ve kalkınmasına katkısı çok önemli olmakla beraber Türkiye'nin yapısal işsizlik, fakirlik, açlık ve dağılım adaletsizliği sorunlarını çözebilecek ilişkilerdir. Bunun yanında Türkiye, AB için çok önemli bir pazardır. Türk ekonomisinin canlılığı ve genç nüfusuyla dinamik yapısı, Türkiye'yi AB ülkeleri için önemli bir pazar konumuna getirmektedir (Keyman, 2009).

Tüm bunlar göz önünde bulundurulduğunda, Türkiye ile AB ilişkileri için yapılacak temel yorum tarafların birbirlerinden kopamayacağıdır (Uysal, 2001). Türkiye'nin aday ülke olarak, AB'nin siyasi kriterlerine uyması ve ekonomik istikrar beklentilerini karşılaması gerekmektedir. Türkiye'nin önünde üzerine düşen sorumlulukları yerine getirmesi gereken bir tam üyelik süreci vardır. Tüm bu kriterlere uyum sağlandığında ve sorumluluklar yerine getirildiğinde Türkiye'nin AB üyeliği süreci hızlanacaktır.

Bu çalışmada klasik kümeleme yaklaşımlarından olan k-medoid kümeleme ve bulanık mantık temelleriyle kümeleme yapan bulanık c-ortalamalar kümeleme yöntemleri kullanılarak OECD üyesi olan 34 ülke sosyoekonomik gelişmişlik açısından analiz edilmiştir. Çalışmada kullanılan veri küresel bir yapılanma gösterdiğinden geleneksel bulanık kümeleme yöntemlerinden bulanık c-ortalamalar yöntemi kullanılmış ve işe yarar sonuçlar üretmiştir. Bununla birlikte benzer bir çalışma, OECD üyeleri dışında yer alan diğer ülkeler eklenerek ve sosyoekonomik gelişmişlik belirten göstergeler zenginleştirilerek, veri yapısına uygun olan başka bulanık kümeleme yöntemleri kullanılarak gerçekleştirilebilir. Böylece daha geniş çerçeveden bakılarak ülkelerin sosyoekonomik gelişmişlikleri mercek altına alınabilir.

KAYNAKLAR DİZİNİ

Abonyi, J., Feil, B., 2007, Cluster Analysis for Data Mining and System Identification, Springer Science & Business Media, p.1-45.

Akın, B., Eren, Ö., 2012, OECD Ülkelerinin Eğitim Göstergelerinin Kümeleme Analizi ve Çok Boyutlu Ölçekleme Analizi ile Karşılaştırma Analizi, Öneri Dergisi, 10 (37), 166-174.

Akses, S., 2010, AB- Güney Kore Serbest Ticaret Anlaşması ve Türkiye' deki Etkisi, http://ikv.org.tr/images/upload/data/files/ikv_degerlendirme_notu_-_ab-guney_kore_serbest_ticaret_anlasmasi_veturkiyedeki_etkisi.pdf, erişim tarihi: 10.06.2015.

Altaş, İ., 1999, Bulanık Mantık: Bulanıklık Kavramı, 3e (Enerji, Elektrik, Elektromekanik) Dergisi, 6, 80-85.

Alvin, R., Christensen, W., 2012, Methods of Multivariate Analysis (3rd Edition), John Wiley& Sons, p.503.

Anderberg, M., 1973, Cluster Analysis for Applications, Academic Press New York.

Aşan, Z., 2007, Kredi Kartı Kullanan Müşterilerin Sosyo-Ekonomik Özelliklerinin Kümeleme Analiziyle İncelenmesi, Dumlupınar Üniversitesi Sosyal Bilimler Dergisi,17, 256-267.

Atalay, A., Tortum, A.,2010, Türkiye'deki İllerin 1997-2006 Yılları Arası Trafik Kazalarına Göre Kümeleme Analizi, Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 16 (3), 335-343.

Avcı, U., 2006, Bulanık Kümeleme Algoritmalarının Karşılaştırmalı Analizi ve Bilgisayar Uygulamaları, Yüksek Lisans Tezi, Ege Üniversitesi Fen Bilimleri Enstitüsü, 78 s. (yayımlanmamış)

Azar, A., El-Said, S., Hassanien, A., 2013, Fuzzy and Hard Clustering Analysis for Thyroid Disease, Computer Methods and Programme in Biomedicine, 111 (1), 1-16.

KAYNAKLAR DİZİNİ (devam ediyor)

Bailey, K., 1975, Sociological Methodology, Wiley, p. 59-128.

Balasko, B., Abonyi, J., Feil, B., 2005, Fuzzy Clustering and Data Analysis Toolbox, Department of Process Engineering University of Veszprem

Cavlak, H., Bostancıoğlu, D., 2015, Avrupa Birliğine Uyum Sürecinde Türkiye'deki Sağlık ve Hasta Hakları, Sosyal Bilimler Dergisi, 2(2), 27-42.

Chen, G., Pham, T., 2001, Fuzzy Sets Fuzzy Logic and Fuzzy Control Systems, CRC Press, p. 1-54.

CIA, 2015, The World Factbook, https://www.cia.gov/library/publications/resources/the-world-factbook/, erişim tarihi: 03.03.2015.

Cormack, R., 1971, A Rewiev of Classification, Journal Of The Royal Statistical Society, 134 (3), 321-367.

Çelik, Ş., 2013, Kümeleme Analizi ile Sağlık Göstergelerine Göre Türkiye' deki İllerin Sınıflandırılması, Doğuş Üniversitesi Dergisi, 14 (2), 175-194.

Çemrek, F., Şentürk, S., Terlemez, L., 2010, Bulanık Kümeleme Analizi ile OECD Ülkelerinin CO_2 Emisyonları Bakımından İncelenmesi, e-Journal of New World Sciences Academy, 5 (3), 52-69.

Doğan, A., Uzun, A., 2014, Serbest Ticaret Anlaşmalarının Türkiye'nin Dış Ticaretine Etkileri, Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi, 15 (1), 325-344.

Döring, C., Lesot, L., Kruse, R., 2006, Data Analysis with Fuzzy Clustering Methods, Computational Statistics & Data Analysis, 51, 192-214.

Elmas, Ç., 2003, Bulanık Mantık Denetleyiciler, Seçkin Yayıncılık, s.51-76.

Erilli, N., Tunç, T., Öner, Y., Yolcu, U., 2008, İllerin Sosyoekonomik Verilere Dayanarak Bulanık Kümeleme Analizi ile Sınıflandırılması, NWSA: Physical Sciences 4 (1), 1-11.

KAYNAKLAR DİZİNİ (devam ediyor)

Eurostat, 2015, Eurostat Database, http://ec.europa.eu/eurostat/data/database, erişim tarihi: 03.03.2015.

Everitt, B., 1993, Cluster Analysis, Hodder Education Publishers, p. 1-88.

Everitt, B., Dunn, G.,1991, Applied Multivariate Data Analysis, London: Edward Arnold, p. 100-126.

Everitt, B., Hothorn, T., 2006, A Handbook of Statistical Analysis Using R, Chapman & Hall/ CRC, p. 245-257.

Giray, S., 2013, Ülkelerin Turizm İstatistikleri Bakımından Farklı Kümeleme Analizi Metotları ile Sınıflandırılması ve Türkiye'nin Bu Oluşumdaki Yeri, Eurosian Economics Association, 695-704, http://avekon.org/papers/816.pdf, erişim tarihi: 20.03.2015.

Giray, S., Gülel, F., 2014, Avrupa Ülkelerinin İntihar Oranlarına Göre Sınıflandırılması, Süleyman Demirel Fen-Edebiyat Fakültesi Sosyal Bilimler Dergisi, 31, 235-247.

Gore, P., 2000, Handbook of Applied Multivariate Statistics and Mathematical Modeling, H. Tinsley, S.D. Brown (Eds.), San Diego: Academic Press, p. 312.

Gökgöz, İ.H., Altınel, F., Gökgöz, P., Koç, İ., 2013, Bankacılık Düzenleme ve Denetleme Kurumu Bankacılık ve Finansal Piyasalar, 7 (2), 13-36.

Grekousis, G., Thomas, H., 2012, Comparison of Two Fuzzy Algorithms in Geodemographic Segmentation Analysis: The Fuzzy C-Means and Gustafson–Kessel Methods, Applied Geography, 34, 125-136.

Gustafson, D.E., Kessel, W.C., 1979, Fuzzy Clustering with a Fuzzy Covariance Matrix, IEEE CDC San Diego, 761-766.

Han, J., Kamber, M., Pei, J., 2006, Data Mining: Concepts and Techniques 2nd Edition, Elsevier, p. 401.

KAYNAKLAR DİZİNİ (devam ediyor)

Han, J., Kamber, M., Pei, J., 2012, Data Mining:Concepts and Techniques 3rd Edition, Elsevier, p. 443-495.

Hartigan, J.A., 1985, Statistical Theory in Clustering, Journal of Classification, 2(1), 63-76.

Hoffman, I., Jarvis, R.,1998, Robust and Efficient Cluster Analysis Using a Shared Near Neighbours Approach, Pattern Recognition Proceedings Fourteenth International Conference, 1, 243-247.

Höppner, F., Klawonn, F., Rudolf, K., Runkler, T., 1999, Fuzzy Cluster Analysis: Methods for Classification Data Analysis and Image Recognition, John Wiley & Sons, p. 5-75.

ILO, 2015, KILM (Key Indicators of Labour Market), http://kilm.ilo.org/2011/Installation/Application2013/kilm13install.htm, erişim tarihi: 05.03.2015.

IMF, 2015, IMF e- Library Data, http://www.imf.org/data, erişim tarihi: 05.03.2015.

Işık, M., Çamurcu, A.Y., 2007, K-means, K-medoids ve Bulanık c-means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 6(11), 31-45.

Jang, J.R., Sun, C.T., Mizutani, E., 1997, Neuro-fuzzy and Soft Computing a Computational Approach to Learning and Machine Intelligence, Prentice Hall, p. 1-42.

Johnson, R., Wichern, D.W., 1998, Applied Multivariate Statistical Analysis Fourth Edition, Prentice Hall, p. 727-799.

Karanfil, S., 1997, Fuzzy Lojik Problemlerinde Üyelik Fonksiyonunun Belirlenmesinde deneysel Verilere Dayanarak Bir Yöntem Geliştirilmesi, Yıldız Teknik Üniversitesi Basım/Yayın Merkezi, s. 9-14.

KAYNAKLAR DİZİNİ (devam ediyor)

Kashigan, S., 1991, Multivariate Statistical Analysis, Radius Press New York, p. 261-270.

Keyman, F., 2009, Avrupa Birliği ve AB İlişkileri Beklentiler ve Kaygılar, O. Esen, F. Başkan (Derl.), Efil Yayınevi, p. 5-21.

Kılıç, İ., Lenger, Ö.F., Bozkurt, Z., 2012, Bulanık Kümeleme Analizi ile Türkiye'deki İllerin Hayvancılık İstatistikleri Bakımından Sınıflandırılması, Kocatepe Veteriner Dergisi, 5 (1), 21-28.

Klir, J.G., Folger, T.,1988, Fuzzy Sets Uncertainty and Information, Prentice Hall, p.1-37.

Klir, G.J., Juan, B., 1997, Fuzzy Set Theory Foundations and Applications, Prentice Hall, p.72-90.

Lange, O., Meyer-Baese, A., Hurdal, M., Foo, S., 2006, A Comparison Between Neural and Fuzzy Cluster Analysis Techniques for Functional MRI, Biomedical Signal Processing and Control, 1, 243–252.

Mamdani, E.H., 1974, Application of Fuzzy Algorithms for Control of Simple Dynamic Plant, Proceedings of the Institution of Electrical Engineers, 121 (12), 1585-1588.

Mamdani, E.H., Assilian, S., 1973, An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller, International Journal of Man-Machine Studies, 7 (1), 1-13.

Manly, B., 1994, Multivariate Statistical Methods, Chapman & Hall, p. 129-145.

Mansoori, E.G., 2011, FRBC: A Fuzzy Rule-Based Clustering Algorithm, IEEE Transactıons on Fuzzy Systems, 19 (5), 960-971.

MFA, 2015, İktisadi İşbirliği ve Gelişme Teşkilatı (OECD), http://www.mfa.gov.tr/iktisadi-isbirligi_ve-gelisme-teskilati-_oecd_.tr.mfa, erişim tarihi: 10.03.2015.

KAYNAKLAR DİZİNİ (devam ediyor)

Mooi, E., Sarstedt, M., 2011, A Concise Guide to Market Research the Process, Data, and Methods Using IBM SPSS Statistics, Springer-Verlag Berlin Heidelberg, p. 237-284.

Naes, T., Mevik, B., 1999, The flexibility of Fuzzy Clustering Illustrated by Examples, Journal of Chemometrics, 13, 435-444.

Nell, T., 2002, Applied Multivariate Analysis, USA: Springer, p.522-523.

OECD, 2015, OECD Database, https://data.oecd.org/, erişim tarihi: 02.03.2015.

Oğuztürk, B.S., 2011, Güney Kore'nin Kalkınmasında İnovasyonun Rolü, Visionary E-Journal / Vizyoner Dergisi, 3 (5), 48-53.

Özdamar, K., 2010, Paket Programlar ile İstatistiksel Veri Analizi 2, Kaan Kitabevi, s. 267-340.

Pang-Ning, T., Steinbach, M., Kumar, V., 2006, Introduction to Data Mining, Boston : Pearson/Addison Wesley, p. 487-568.

Sharma, S., 1996, Applied Multivarite Techniques, John Wiley & Sons, p.185-229.

Sintas, F., Cadenas, J., Martin, F., 1999, Membership Functions in the Fuzzy c-means Algorithm, Fuzzy Sets and Systems, 101 (1), 49-58.

Sönmez, H., Er, F., 2006, Türkiye' de İllere Göre İç Göç Hareketlerinin Modern Kümeleme Teknikleri ile İncelenmesi, Eskişehir Osmangazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 20 (1), 17-32.

Struyf, A., Hubert, M., Rousseeuw, P., 1997, Clustering in an Object-Oriented Environment, Journal of Statistical Software, 1 (4), 1-30.

Şahin, M., Hamarat, B., 2002, G10-Avrupa Birliği ve OECD Ülkelerinin Sosyoekonomik Benzerliklerinin Fuzzy Kümeleme Analizi ile Belirlenmesi, erc/ODTÜ Uluslararası Ekonomi Kongresi, 6, 11-14.

KAYNAKLAR DİZİNİ (devam ediyor)

Şentürk, S., 2006, Deney Planlamasında Bulanık Mantık Yaklaşımı, Eskişehir : Anadolu Üniversitesi, s. 24-27.

Tatlıdil, H., 2002, Uygulamalı Çok Değişkenli İstatistiksel Analiz, Akademi Matbaası Ankara, s.330.

Terlemez, L., 2001, Kümeleme Analizi ile Avrupa Birliğine Aday Ülkelerin Ekonomik Durumlarının İncelenmesi Yüksek Lisans Tezi, Anadolu Üniversitesi, 71 s.

Tinsley, H., Brown, S., 2000, Handbook of Applied Multivariate Statistics and Mathematical Modeling, H. Tinsley, S.D. Brown (Eds.), San Diego: Academic Press, p. 3-5.

Türkşen, B., 2005, An Ontological and Epistemological Perspective of Fuzzy Set Theory, Amsterdam ; Oxford : Elsevier, p.123-143.

UN, 2015, UNdata, https://data.un.org/, erişim tarihi:03.03.2015.

Uysal, C., 2001, Türkiye-Avrupa Birliği İlişkilerinin Tarihsel Süreci ve Son Gelişmeler, Akdeniz İ.İ.B.F. Dergisi, 1, 140-153.

Wang, W., Zhang, Y., 2007, On Fuzzy Cluster Validity İndices, Fuzzy Sets and Systems, 158, 2095-2117.

Wilks, D., 2011, Statistical Methods in the Atmospheric Sciences, Elsevier, p. 603-604.

Worldbank, 2015, Worldbank Data, http://data.worldbank.org/, erişim tarihi: 03.03.2015.

Xu, R., Wunsch, D., 2005, Survey of Clustering Algorithms, Neural Networks, IEEE Transactions, 16 (3), 645-678.

Yang, M.S., 1993, A Survey of Fuzzy Clustering, Mathematical and Computing Modelling, 18 (11), 1-16.

KAYNAKLAR DİZİNİ (devam ediyor)

Yeloğlu, H.O., 2009, Bilgi Ekonomisi Değişkenlerine Yönelik İlk İzlenimler: Türkiye-OECD Ülkeleri Karşılaştırmaları (1995-1999), Bilgi Dünyası, 10 (2), 245-260.

Yılancı, V., 2010, Bulanık Kümeleme Analizi İle Türkiye'deki İllerin Sosyoekonomik Açıdan Sınıflandırılması, Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 15 (3), 453-470.

Yıldız, Z., 1989, Banka Müşterilerinin Demografik ve Sosyo-Ekonomik Özellikler Bakımından Gruplandırılmasında Kümeleme Çözümlemesi ve Bir Uygulama Yüksek Lisans Tezi, Anadolu Üniversitesi, 42 s.

Zadeh, L.A., 1965, Fuzzy Sets, Information and Control, 8, 338-353.