Mühendislik Bilimleri ve Tasarım Dergisi
8(5), 218-229, 2020
e-ISSN: 1308-6693
Araştırma Makalesi

Journal of Engineering Sciences and Design
DOI: 10.21923/jesd.828051
Research Article

Özel Sayı: Uluslararası Mühendislikte Yapay Zeka ve Uygulamalı Matematik Konferansı (UMYMK 2020)

Special Issue: International Conference on Artificial Intelligence and Applied Mathematics in Engineering (ICAIAME 2020)

MODELE DAYALI KÜMELEME ANALİZİNDE OPTİMUM KÜMELEME İÇİN YENİ BİR YAKLAŞIM

Serkan Akogul[1*], Maruf Gogebakan[2]

[1] Pamukkale Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Denizli, Türkiye
[2] Bandırma Onyedi Eylül Üniversitesi, Denizcilik Fakültesi, Denizcilik İşletmeleri Yönetimi Bölümü Balıkesir, Türkiye

Anahtar Kelimeler

Bilgi Kriterleri,
Değişken Bileşenlere Göre Modele Dayalı Kümeleme,
Normal Karma Modeller,
TOPSIS.

Öz

Sonlu karma modellerde bileşen (küme) sayısının belirlenmesi önemli bir problem olup normal karma modeller, sonlu karma dağılımlarda sıklıkla kullanılmaktadır. Bu çalışmada, çok kriterli karar verme yöntemlerinden biri olan TOPSIS yöntemi ile çok değişkenli veri setinin modellenmesinde yeni bir kümeleme yöntemi önerilmiştir. Önerilen yöntemde, çok değişkenli verinin her bir değişkeni tek değişkenli normal karma dağılımlarla modellenip, bileşen sayısına göre elde edilen bilgi kriteri değerleri kullanılarak bir karar matrisi oluşturulmuştur. Karar matrisi kullanılarak TOPSIS yöntemi ile değişkenlerdeki bileşen sayısı belirlenmiştir. Bileşen bulunmayan homojen değişkenler elenerek boyut indirgenmiş olup heterojen değişkenlerdeki bileşen sayılarına göre oluşabilecek karma modeller için alternatif bileşen sayıları hesaplanmıştır. Alternatif bileşen sayıları içerisinden en uygun bileşen sayısı ve uygun karma model yine TOPSIS yöntemi ile belirlenmiştir. Böylece çok değişkenli veride boyut indirgeme ve değişken seçimi ile küme sayısı tahmini yapılmıştır. Önerilen yaklaşımın başarısı gerçek veri seti üzerinde test edilmiş olup veri setinin küme sayısı doğru olarak belirlenmiştir. Ayrıca bu yaklaşım, gözlemlerin sınıflandırma başarısını da arttırmıştır.

A NEW APPROACH TO OPTIMUM CLUSTERING IN MODEL-BASED CLUSTER ANALYSIS

Keywords

Information Criteria,
Model-Based Clustering According to Variable Components,
Gaussian Mixture Models,
TOPSIS.

Abstract

Determining the number of components in finite mixture models is an important problem, and normal mixture models are frequently used in finite mixture distributions. In this study, In this study, a new clustering method is proposed for modeling multivariate data set with TOPSIS method. In the proposed method, each variable of multivariate data is modeled with univariate normal mixture distributions and a decision matrix is created by using the information criterion values obtained according to the number of components. The number of components in the variables was determined with TOPSIS method using the decision matrix. Homogeneous variables without components are eliminated and the size is reduced, and alternative component numbers are calculated for mixture models that can be formed according to the number of components in heterogeneous variables. Among the alternative number of components, the most suitable component number and suitable mixture model were determined by TOPSIS method. Thus, in multivariate data, number of clusters were estimated with dimension reduction and variable selection. The success of the proposed approach has been tested on the real dataset and the number of clusters of the data set has been determined correctly. In addition, this approach increased the classification success of the observations.

* İlgili yazar / Corresponding author: sakogul@pau.edu.tr, +90-258-296-3695

Alıntı / Cite		
Akogul, S., Gogebakan, M. (2020). Modele Dayalı Kümeleme Analizinde Optimum Kümeleme İçin Yeni Bir Yaklaşım, Mühendislik Bilimleri ve Tasarım Dergisi, 8(5), 218-229.		
Yazar Kimliği / Author ID (ORCID Number)	**Makale Süreci / Article Process**	
S. Akogul, 0000-0002-0346-4308	**Başvuru Tarihi / Submission Date**	19.11.2020
M. Gogebakan, 0000-0003-0447-8311	**Revizyon Tarihi / Revision Date**	03.12.2020
	Kabul Tarihi / Accepted Date	09.12.2020
	Yayım Tarihi / Published Date	29.12.2020

1. Giriş (Introduction)

Modele dayalı kümeleme p-boyutlu çok değişkenli heterojen veriyi anlamlı alt gruplara bölmek için kullanılan yöntemlerden biridir (Fraley ve Raftery, 2002). Sonlu karma dağılımların modele dayalı kümelenmesi, kümeleme analizinde sık kullanılan bir yaklaşımdır (McLachlan ve Peel, 2004; Fop ve Murphy, 2018). Sonlu karma modellerin kümelenmesinde veri setinin kaç bileşenle temsil edileceği halen tartışılan önemli bir problemdir (Bozdogan, 1994; Oliveira-Brochado ve Martins, 2005; Nguyen ve McLachlan, 2015).

Çok değişkenli normal karma dağılımların her bir değişkenindeki bileşen çok değişkenli heterojen verideki bir kümeye karşılık gelir (McLachlan ve Chang, 2004). Verideki değişkenlerin bağımsız altkümelerinden çoklu kümelenme yapılarını incelemek için modele dayalı bir yöntem kullanılır. Çok değişkenli verilerde oluşan çoklu kümelerin sayısını ve yapısını tanımlamada değişkenlerin parçalanmaları kullanılır (Galimberti ve Soffritti, 2007; Soffritti, 2003). Seo ve Kim (2012) "Normal karma modellerde kök seçimi" başlıklı çalışmalarında normal dağılımların karma modelinde verideki alt grup yapılarını tanımlamak için kök seçimi olarak adlandırdıkları bileşen sayısı tahmini metodunu geliştirdiler. Modele dayalı kümelenmede küme sayısının belirlenmesinde bileşen seçimi algoritmasının verimli çalışması için maksimum küme sayılarının belirlenmesi önemli bir problemdir (Fraley ve Raftery, 1998).

Çok değişkenli verilerde değişkenlerin bileşenlerine dayalı küme sayısının belirlenmesi için minimum ve maksimum küme sayıları arasından en uygun küme sayısının belirlenmesi için değişken veri parçalanması metodu kullanılır (Gogebakan ve Erol, 2019). Çok değişkenli verilerde modele dayalı küme sayısı tahmini bilgi kriterleri kullanılarak optimizasyon ile elde edilir (Mirzal, 2020). Bilgi kriterleri verinin parametre tahmininden elde edilen verilere dayalı hata payı içerdiğinden farklı sonuçlar çıkabilir (Akogul ve Erisoglu, 2016). Akogul ve Erisoglu (2017) çalışmalarında küme sayısının belirlenmesinde bilgi kriterlerinden kaynaklı hatayı en aza indirgemek için çok kriterli karar verme (ÇKKV) yöntemlerinden Analytic Hierarchy Process (AHP)'yi kullanmışlardır. Karar matrisindeki ölçütler bilgi kriterleri ve seçeneklerde küme sayıları olacak şekilde problemin çözümüne ulaşmışlardır.

Bu çalışmada, modele dayalı kümelemede küme sayısı tahmini için yeni bir yöntem önerilmiştir. Önerilen yöntem ile çok değişkenli verinin her bir değişkeni tek değişkenli normal karma dağılımlarla modellenmiştir. Değişkenlerdeki bileşen sayısı, bilgi kriteri ile oluşturulan karar matrisi yardımıyla TOPSIS yöntemi kullanılarak belirlenmiştir. Bileşen sayısı bir olan homojen değişkenlerin, modellerin oluşturulmasında küme sayısını etkilemediği için boyut indirgenmiş ve değişken seçim metodu önerilmiştir. Böylelikle daha az değişkenle veri setinin küme sayısı tahmin edilmiştir. Önerilen yaklaşım, gerçek veri seti üzerinde test edilmiş olup daha az değişkenle doğru küme sayısı ve yüksek sınıflama başarısı elde edilmiştir.

Çalışmanın ikinci bölümünde konu ile alakalı kaynak araştırmasına yer verilmiştir. Üçüncü bölümde metot ve yöntemler açıklanmış olup önerilen yöntem ortaya konulmuştur. Dördüncü bölümde önerilen yöntem gerçek veri seti üzerinde test edilmiş ve hesaplamalara yer yerilmiştir. Beşinci bölümde de sonuçlara ve önerilere yer verilmiştir.

2. Kaynak Araştırması (Literature Survey)

Karma dağılım modeliyle ilgili yapılan ilk çalışma, Pearson (1894) tarafından aynı coğrafik bölgede yaşayan farklı yengeç türlerini modellemek için yaptığı çalışma kabul edilmektedir. Bu çalışmada, iki tek değişkenli normal dağılımın karmasını kullanarak yengeç türleri modellenmiş ve momentler yöntemi ile parametre tahminleri yapılmıştır.

Day (1969) ve Binder (1978) ilk kez çok değişkenli normal dağılımların karması üzerine çalışmışlar ve parametre tahminlerinde bulunmuşlardır.

Dempster ve ark. (1977) tamamlanmamış veri setinden en çok olabilirlik tahminleri hesaplamak için Expectation–Maximization (EM) algoritmasını önermişlerdir.

McLachlan ve Peel (2004)'in "Finite Mixture Models" başlıklı kitabı hem teorik hem de uygulama açısından sonlu karma modeller için temel ve kapsamlı bir çalışmadır. EM algoritması kullanılarak sonlu karma dağılım modelindeki parametre tahminleri, en çok olabilirlik tahmin edicilerinin özellikleri, karma modeldeki bileşen sayısının belirlenmesi gibi konular bu kitapta ele alınan başlıca hususlardır.

Servi (2009) tez çalışmasında, çok değişkenli normal dağılımların karmasına dayalı kümeleme analizinde model bileşen sayısı tahmini için bir yöntem önererek uzaktan algılanmış çok bantlı uydu görüntü verisini sınıflandırmıştır.

Gogebakan (2017) "Karma Dağılım Modelleri Kullanılarak Çok Değişkenli Veride Grup Yapılarının Belirlenmesi, Ayrıştırılması, Kümelenmesi ve Sınıflandırılması" başlıklı doktora tezinde, bileşenlere dayalı karma dağılım yaklaşımı ve normal dağılımların karmasına dayalı kümeleme analizi üzerine çalışmalar yapmıştır.

Akogul (2018), "Çok Değişkenli Verilerin Modele Dayalı Kümeleme Analizinde Kümeleme Etkinliğinin Arttırılması" başlıklı doktora tezinde karma dağılım, bilgi kriterleri ve ÇKKV yöntemi ile küme sayısının belirlenmesi konuları üzerine çalışmalar yapmıştır.

Hwang ve Yoon (1981) tarafından ÇKKV yöntemi çerçevesinde ideal çözüme en yakın ideal olmayan çözüme ise en uzak sonucu verecek bir yöntem olan TOSIS geliştirilmiştir.

Wang ve Lee (2009) çalışmalarında, alternatifleri çeşitli kriterler altında sistematik olarak değerlendirmek için, öznel ve nesnel ağırlıklar kullanarak yeni bir bulanık TOPSIS önermişlerdir.

Ishizaka ve Nemery (2013) tarafından yazılan bu kitap, ÇKKV yöntemlerine bir giriş ve ardından bu alanda kullanılan önde gelen yöntemlerin her biri hakkında daha ayrıntılı bölümler sunmaktadır.

Burak vd. (2015) çalışmalarında, alternatifler arasından ergonomik olarak uygun olan telefon seçimi için TOPSIS metodunun sezgisel bulanık ortama genişletilmesi önerilmiştir.

Yıldırım ve Önder (2015) kitabında, karar teorisi, çok kriterli karar verme sürecinden ve bazı önemli ÇKKV yöntemlerini uygulamalarıyla birlikte anlatmışlardır.

Şahin ve Supçiller (2015) çalışmalarında, tedarikçi seçimi ve değerlendirme probleminin çözümü için Analitik Hiyerarşi Prosesi (AHP), TOPSIS ve k-ortalamalar yöntemlerinin kullanıldığı bir karar destek sistemi önerilmiştir.

Akalp ve Özok (2017) çalışmalarında sözel belirsizliklerin bulunduğu bulanık ortamlarda, özellikle İş Sağlığı ve Güvenliği çalışmalarında çalışanlar açısından daha fazla riskli olan istasyonun seçimi amacıyla Bulanık TOPSIS algoritmasının kullanılabilirliğinin araştırmışlardır.

Akogul ve Erisoglu (2017) çalışmalarında küme sayısının belirlenmesinde bilgi kriterlerinden kaynaklı hatayı en aza indirgemek için ÇKKV yöntemlerinden AHP'yi kullanmışlardır. Böylece küme sayısı belirleme problemi bir ÇKKV problemi olarak ele alınmıştır.

Acer ve Kalender (2020) antrepoların performansının değerlendirilmesinde önemli olan ana kriterler belirlenmeye çalışılmış ve kriter ağırlıklarının hesaplanmasında entropiye dayalı TOPSIS yöntemi ile antrepolar sıralanmıştır.

Çetin ve Alvatı (2020) çalışmalarında, boji malzemesi için S235, S275 ve S355 malzemeler dikkate alınarak en uygun karbonlu çelik alaşımının belirlenmesi amaçlanmıştır. Modelin çözümü için çok kriterli karar verme yöntemlerinden olan TOPSIS ve VIKOR algoritmaları kullanılmıştır.

3. Materyal ve Yöntem (Material and Method)

Çok değişkenli veri setlerinin kümelenmesinde normal dağılımların karmasına dayalı kümeleme yaygın olarak kullanılmaktadır. Çalışmada ilk önce çok değişkenli veri setinin her bir değişkeni Tek Değişkenli Normal Karma Model (TDNKM) ile modellenip her bir değişkendeki bileşen sayısı tahmin edilmiştir. Bileşen sayısı tahmini için bilgi kriterlerinden yararlanılmıştır. Lakin birçok bilgi kriteri aynı veri seti için farklı bileşen sayısı tahminleri vermektedir. Bu sorunun üstesinden gelebilmek içinde bilgi kriterlerinden elde edilen karar matrisi kullanılarak

TOPSIS yöntemi ile uygun bileşen sayısı belirlenmiştir. Her bir değişkenden elde edilen bileşen sayılarına göre tam veri seti, tek bileşene sahip değişkenlerden arındırılıp daha az değişkene sahip indirgenmiş veriye dönüştürülmüştür. İndirgenmiş veri seti için minimum ve maksimum bileşen sayısı değişken veri parçalanması metodu ile belirlenmiştir. Bu sayılara göre indirgenmiş veri tek ya da çok değişkenli normal karma model ile modellenmiş ve yine bilgi kriterlerinden oluşan karar matrisi kullanılarak TOPSIS ile veri setinin bileşen sayısı belirlenmiştir. Ayrıca karşılaştırma yapabilmek için tam ve indirgenmiş veri setleri, tahmin edilen bileşen sayısı kullanılarak normal karma dağılımlar ile modellenmiş ve gerçek üyelikleri bilinen gözlemlerin ait oldukları kümeye atanma oranları olan doğru sınıflandırma oranı (DSO) hesaplanmıştır. Önerilen yöntem küme sayısı tahmininde ve DSO göre daha iyi sonuçlar vermiştir.

3.1. Tek Değişkenli Normal Karma Modeller (Univariate Normal Mixture Models)

Tek değişkenli normal karma dağılımlar

$$f(x;\theta)=\sum_{j=1}^{k}\pi_j f_j\left(x;\mu_j,\sigma_j\right) \quad (1)$$

şeklinde gösterilir. Burada k tek değişkenli dağılımın karma bileşen sayısı, π_j karma olasılık ağırlıklarını temsil etmektedir. Tek değişkenli normal karma dağılımların olasılık yoğunluk fonksiyonları

$$f(x;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}, x\in\mathbb{R} \quad (2)$$

şeklinde ifade edilir. Burada değişkenin gözlem değerlerine dayalı ortalama vektörü μ ve standart sapması σ olarak elde edilir. Tek değişkenli karma modellerin parametreleri Beklenti (E) ve En büyük (M) yapma algoritması (EM) kullanılarak elde edilir. z tamamlanmış tek değişkenli karma dağılımlarda etiket vektörü olmak üzere $\{X_1,X_2,\dots,X_n,Z_1,Z_2,\dots,Z_n\}$ verisindeki log-olabilirlik fonksiyonu,

$$logL(\pi,\theta;x_1,x_2,\dots,x_n,z_1,z_2,\dots z_n)=\sum_{i=1}^{n}\sum_{j=1}^{k}z_{ji}log\left[\pi_j f\left(x;\theta_j\right)\right] \quad (3)$$

olarak hesaplanır. EM algoritmasında E ve M adımları aşağıdaki şekilde uygulanır.:

- ***E Adımında***: Etiket vektörü z_{ji} değerlerini tahmin etmek için koşullu beklenen değer,

$$\hat{z}_{ji}=E\left(z_{ji}\middle|x;\pi,\theta\right)=\frac{\pi_j f\left(x;\theta_j\right)}{\sum_{j=1}^{k}f\left(x;\theta_j\right)} \quad (4)$$

şeklinde tahmin edilir.

- ***M Adımında:*** Tek değişkenli karma olasılık ağırlıkları π_j ve $\sum_{j=1}^{k}\pi_j=1$ olmak üzere log-olabilirlik fonksiyonunu maksimize etmek için,

$$logL(\pi,\theta;x_1,x_2,\dots,x_n,z_1,z_2,\dots z_n)=\sum_{i=1}^{n}\sum_{j=1}^{k}\hat{z}_{ji}log\left[\pi_j f\left(x;\theta_j\right)\right] \quad (5)$$

eşitliği elde edilir. Tek değişkenli normal karma dağılımlarda, EM algoritması log-olabilirlik fonksiyonundaki parametre değerleri değişmeyene kadar çalışmaya devam eder (Ridolfi ve Idier, 2001; McLachlan ve Peel, 2004).

3.2. Normal Karma Modeller (Gaussian Mixture Model-GMM)

Normal karma modeller

$$f(x_i;\theta)=\sum_{j=1}^{k}\pi_j f_j\left(x_i;\psi_j\right) \quad (6)$$

şeklinde ifade edilir. Burada $0 < \pi_j < 1$ ve $\sum_{i=1}^{k} \pi_j = 1$ olacak şekilde π_j ler j. kümenin karma ağırlık oranını, $\psi_j = (\mu_j, \Sigma_j)$ parametrelerin vektörünü temsil etmektedir. $f_j(x_i; \mu_j, \Sigma_j)$ çok değişkenli Normal dağılımların olasılık yoğunluk fonksiyonu

$$f_j(x_i; \mu_j, \Sigma_j) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_j|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)\right\} \quad (7)$$

şeklinde ifade edilir. Burada Eşitlik (6)'daki $\theta_j = \pi_1, \dots, \pi_k,\ \psi_1, \dots, \psi_k$ vektörleri $\mathbf{\Psi}$ uzayında normal karma dağılımların çok değişkenli parametre vektördür.

Normal karma modeller için olasılık yoğunluk fonksiyonlarından elde edilen olabilirlik fonksiyonu ve buna bağlı hesaplanan log-olabilirlik fonksiyonu

$$logL(\mathbf{\Psi}) = \sum_{i=1}^{n} log\left(\sum_{j=1}^{k} \pi_j f_j(x_i; \theta_j)\right) \quad (8)$$

olarak elde edilir. Gözlenen $\boldsymbol{x}_i$'nin karma modelin j'inci bileşenine ait olma olasılığı

$$\tau_j(\boldsymbol{x}_i; \boldsymbol{\Psi}) = \pi_j \Phi_j(\boldsymbol{x}_i; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \Big/ \sum_{m=1}^{k} \pi_m \Phi_j(\boldsymbol{x}_i; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \quad (9)$$

olup $\boldsymbol{\Psi}$ parametresinin EM algoritmasıyla hesaplanan en çok olabilirlik tahminleri

$$\pi_j^{(m+1)} = \sum_{i=1}^{n} \frac{\tau_j(\boldsymbol{y}_i; \boldsymbol{\Psi}^{(m)})}{n} \quad (10)$$

$$\boldsymbol{\mu}_j^{(m+1)} = \frac{1}{\sum_{i=1}^{n} \tau_j(\boldsymbol{x}_i; \boldsymbol{\Psi}^{(m)})} \sum_{i=1}^{n} \tau_j(\boldsymbol{x}_i; \boldsymbol{\Psi}^{(m)}) \boldsymbol{x}_i \quad (11)$$

$$\boldsymbol{\Sigma}_j^{(m+1)} = \frac{1}{\sum_{i=1}^{n} \tau_j(\boldsymbol{x}_i; \boldsymbol{\Psi}^{(m)})} \sum_{i=1}^{n} \tau_j(\boldsymbol{x}_i; \boldsymbol{\Psi}^{(m)})(\boldsymbol{x}_i - \boldsymbol{\mu}_j^{(m+1)})(\boldsymbol{x}_i - \boldsymbol{\mu}_j^{(m+1)})^T \quad (12)$$

iterasyonları ile elde edilir (McLachlan ve Peel, 2004; Akogul, 2018; Andriyanov vd., 2020).

3.3. Normal Karma Modellerde Bileşen Sayısının Belirlenmesi ve Değişkenlerdeki Parçalanmaya Göre Aday Bileşen Sayısının Hesaplanması (Determining the Number of Components in Gaussian Mixture Models and Calculating the Number of Candidate Components According to the Fragmentation in the Variables)

Normal karma modellerde önemli problemlerden birisi uygun bileşen (küme) sayısının belirlenmesidir. Uygun bileşen sayısı, genellikle normal karma dağılımların log-olabilirlik değerinden hesaplanan bilgi kriterleri yardımıyla belirlenir. Çalışmada literatürde yaygın olarak kullanılan Akaike bilgi kriteri (AIC) (Akaike, 1974), Bilgi Kriterinin Küçük Örneklem Yanlı Düzeltilmiş Hali (AICc) (Hurvich ve Tsai, 1989) ve Bayesci bilgi kriteri (BIC) (Schwarz, 1978) kullanılmıştır.

Normal karma modellerde bileşen sayısı tahmini için bilgi kriterleri,

$$-2logL(\mathbf{\Psi}) + C \quad (13)$$

şekilde ifade edilir. Burada C, ceza terimi olarak adlandırılır ve modeldeki serbest parametre sayısı (d) veya veri setindeki gözlem sayısına (n) bağlı olarak belirlenir (McLachlan ve Peel, 2004).

Karma modellerde matrisinde k modeldeki bileşen ve p değişken sayısı olmak üzere serbest parametre sayısı $d = (k-1) + (kp) + kp\left(\frac{k+1}{2}\right)$ ile hesaplanır. Eşitlik (13) de verilen bilgi kriteri için j bileşen sayısını göstermek üzere $Bilgi\ Kriteri(j) \leq Bilgi\ Kriteri(j+1)$ koşulunu sağlayan ilk j sayısı, uygun küme sayısı (k) olarak seçilir.

Çalışmada kullanılan bilgi kriterleri

$$\mathrm{AIC} = -2logL(\widehat{\boldsymbol{\Psi}}) + 2d \quad (14)$$

$$\mathrm{AIC_c} = -2logL(\widehat{\boldsymbol{\Psi}}) + 2d(n/(n-d-1)) \quad (15)$$

$$\mathrm{BIC} = -2logL(\widehat{\boldsymbol{\Psi}}) + dlog(n) \quad (16)$$

eşitlikleriyle hesaplanır (Akogul ve Erisoglu, 2016).

Çok değişkenli veri setinin her bir değişkenindeki parçalanmaya göre veriyi modelleyecek karma modellerde oluşabilecek aday bileşen ve karma model sayılarının hesaplanması için Gogebakan ve Servi (2019) tarafından önerilen yaklaşım kullanılmıştır. Çok değişkenli veri setinde p değişken sayısı, $k_j, j = 1, \dots, p$ olacak şekilde her bir değişkendeki karma modelde oluşabilecek alternatif minimum ve maksimum küme sayıları $k_{min} = max\{k_j\}$ ve $k_{max} = \prod_{j=1}^{p} k_j$ denklemleri ile hesaplanır. Değişkenlerdeki bileşenlerin oluşturduğu küme sayılarına göre toplam karma modellerin sayısı $M_{Toplam} = 2^{\prod_{j=1}^{p} k_j} - 1$ eşitliği ile elde edilir (Gogebakan ve Erol, 2019).

3.4. TOPSIS Yöntemi (TOPSIS Method)

ÇKKV, birden fazla kriterin optimizasyonu ile çözüm kümeleri içerisinden en iyi alternatifin belirlendiği yöntemlerdir (Yıldırım ve Önder, 2015). Veri setinin modellenmesinde, normal karma modeldeki bileşen sayısını belirlemek için çok sayıda bilgi kriterinin optimize edilmesi gerekmektedir. Modeldeki bileşen sayısını belirleme ÇKKV problemi olarak ele alınabilir (Ishizaka ve Nemery, 2013; Akogul ve Erisoglu, 2017; Khan ve Young, 2018).

Hwang ve Yoon (1981) tarafından ÇKKV yöntemi çerçevesinde ideal çözüme en yakın ideal olmayan çözüme ise en uzak sonucu verecek bir yöntem olan TOSIS geliştirilmiştir. TOPSIS yöntemi uygulamadaki kolaylık açısından son zamanlarda literatürde sıkça uygulamaları bulunmaktadır. TOPSIS yöntemi için altı aşamalı bir algoritma kullanılmaktadır (Wang ve Lee, 2009; Perçin ve Sönmez, 2018; Ece, 2019; Özgüner, 2020; Acer ve Kalender, 2020):

- ***Karar matrisi***: $D = [d_{ij}]_{mxn}$ karar matrisi, mxn boyutlu, n adet belirleme kriteri ve m adet bu kriterlerin alternatifinden oluşur. Burada d_{ij} i'inci alternatifin $(i = 1, \dots, m)$ j'inci belirleme kriterine $(j = 1, \dots, k)$ göre önem seviyesini göstermektedir.

- ***Normalize matrisi:*** $N = [n_{ij}]_{mxn}$ normalize matrisi, D karar matrisinin elemanlarından $n_{ij} = d_{ij} / \sqrt{\sum_{i=1}^{m} d_{ij}^2}$ eşitliği ile oluşturulur.

- ***Ağırlıklandırılmış Normalize matrisi****:* Normalize matrisin her bir elemanı $\sum_{j=1}^{n} w_j = 1$ olmak üzere w_j gibi bir olasılık ağırlığı ile çarpılarak $(v_{ij} = w_j n_{ij})$ ağırlıklandırılmış normalize matris $V = [v_{ij}]_{mxk}$ elde edilir.

- ***İdeal ve negatif ideal çözümler****:* Ağırlıklandırılmış normalize matrisin (V) sütunlarındaki belirleme kriterlerinden elde edilen değerler optimizasyon için kullanılır. Matristeki elemanlar ile minimizasyon için sütunlardaki minimum değerli $A^+ = \{v_1^+, v_2^+, \dots, v_k^+\}$ elemanlar ideal çözümleri, maksimum değerli $A^- = \{v_1^-, v_2^-, \dots, v_k^-\}$ elemanlar ise negatif ideal çözümleri oluşturacak şekilde belirlenir. Eğer problemin çözümünde optimizasyonda maksimizasyon yaklaşımı kullanılacaksa ideal ve negatif ideal elemanları tam tersi olarak belirlenecektir.

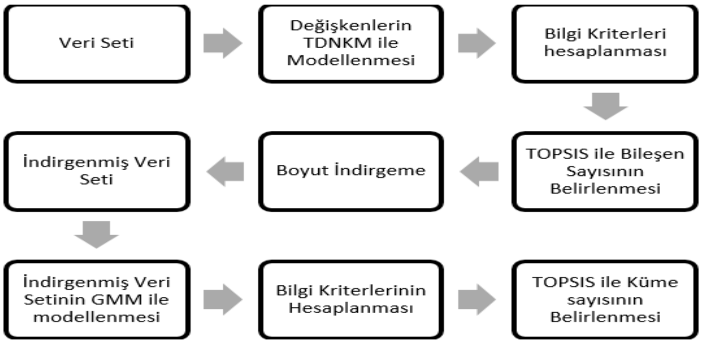
- ***İdeal ve negatif ideal ayırım ölçülerinin hesaplanması****:* İdeal ve negatif ideal noktalara olan uzaklıklar Öklid uzaklığı kullanılarak elde edilir. S_i^+ ve S_i^- ölçülerinin sayısı alternatiflerin sayısı kadar olmak koşuluyla $S_i^+ = \sqrt{\sum_{j=1}^{k} (v_{ij} - v_j^+)^2}$ ideal ayrım ve $S_i^- = \sqrt{\sum_{j=1}^{k} (v_{ij} - v_j^-)^2}$ negatif ideal ayrım ölçüleri olacak şekilde hesaplanır.

- ***İdeal çözüme göreli yakınlığın hesaplanması:*** İdeal ve negatif ideal ayırım ölçüleri kullanılarak alternatifler için ideal çözüme göreli yakınlık değeri $C_i^+ = S_i^- / (S_i^- + S_i^+)$ olacak şeklinde hesaplanır. Burada $0 \leq C_i^+ \leq 1$ olacak şekilde $C_i^+ = 1$ alternatifin ideal çözüme, $C_i^+ = 0$ alternatifin negatif ideal çözüme mutlak yakınlığını gösterir.

3.5. Değişkenlerdeki Bileşen Sayısının TOPSIS Yöntemiyle Belirlenmesi ve Veri Setinin Normal Karma Model ile Kümelenmesi (Determining the Number of Components in Variables by the TOPSIS Method and Clustering the Data Set with Gaussian Mixture Model)

Çok değişkenli verilerde değişken sayısının fazla olması kümelemeyi olumsuz etkilemektedir. Modele dayalı kümelemede de değişken sayısının artmasıyla bileşen sayısının belirlenmesi ve parametre tahminlerinde zorluklar yaşanmaktadır. Bundan dolayı uygun değişkenlerin belirlenmesi önem arz etmektedir.

Çalışmada çok değişkenli veri setinin her bir değişkeni tek değişkenli normal karma modellerle modellenmiş ve değişkenlerdeki parçalanmalar yani bileşenler hesaplanmıştır. Parçalanmanın olmadığı yani tek bileşene sahip değişkenler veri setinden ayıklanarak değişken seçim yöntemi uygulanmıştır. Böylelikle kümelemeye etkisi olmadığı düşünülen değişkenler çıkarılıp daha az değişkene sahip indirgenmiş veri elde edilmiştir. Her bir değişkendeki bileşen sayıları, bilgi kriterlerinin oluşturduğu karar matrisi kullanılarak TOPSIS yöntemiyle belirlenmiştir. Daha sonra indirgenmiş veri seti için bileşenlerdeki parçalanma sayılarına göre minimum ve maksimum aday bileşen sayısı hesaplanmıştır. Aday bileşen sayılarına göre indirgenmiş veri seti normal karma modelle kümelenmiş ve bilgi kriterleri hesaplanmıştır. TOPSIS yönteminde kullanılacak olan karar matrisinde, aday bileşen sayıları alternatifleri ve bilgi kriterleri de ölçütleri oluşturmaktadır. Böylelikle veri seti için en uygun bileşen sayısı TOPSIS yöntemiyle belirlenmiş olur. Önerilen bu yeni yaklaşım Şekil 1'de gösterilmiştir.



Şekil 1. Küme sayısının belirlenmesi için önerilen yaklaşım (The suggested approach for determining the number of clusters)

Uygun bileşen sayısının belirlenmesinden sonra veri seti normal karma modele kümelenip, her bir gözlemin oluşan kümelere atanma olasılıkları ve küme üyelikleri belirlenir. Böylelikle gerçek üyelikleri bilinen gözlemlerin doğru sınıflama başarısı hesaplanmış olur.

4. Deneysel Sonuçlar (Experimental Results)

Bu çalışmada çok değişkenli verilerin normal karma modellere dayalı kümelenmesi için küme sayısının belirlenmesine yönelik değişken bileşenlerine dayalı yeni bir kümeleme yöntemi öne sürülmüştür. Çalışmadaki yöntem çok değişkenli İris veri seti üzerinde uygulanmıştır. Değişkenlerdeki bileşen sayısını belirlemek için değişkenler tek değişkenli normal karma modeller ile bileşen sayılarına göre AIC, AICc ve BIC bilgi kriterleri elde edilmiştir. Alternatifler bileşen sayısı ölçütlerde bilgi kriterleri olan karar matrisleri hesaplanmıştır. Yöntem tek bir değişken için detaylı anlatılmış olup diğer değişkenler için sonuçlar verilmiştir.

Fisher tarafından 1936 yılında sunulan ve kümelemede yaygın olarak kullanılan Iris veri seti, 3 küme (setosa, virginica ve versicolor), 4 değişken (Sepal length, Sepal width, petal length ve petal width) ve 150 gözleme sahip çok değişkenli bir veri setidir. İris veri setinin birinci değişkeni olan Sepal length için bileşen sayılarına göre hesaplanın bilgi kriterlerinin oluşturduğu karar matrisi Tablo 1 de verilmiştir.

Tablo 1. Sepal length için bileşen sayısına göre bilgi kriterleri değerlerinin oluşturduğu karar matrisi (Decision matrix formed by the information criteria values according to the number of components for the sepal length)

		Ölçütler		
	Bileşen sayısı	**AIC**	**AICc**	**BIC**
Alternatifler	**1**	372.08	372.16	378.10
	2	365.72	366.14	380.77
	3	366.95	367.54	385.02
	4	366.05	367.07	390.14
	5	370.11	371.69	400.21

Karar matrisi yardımıyla elde edilen normalize edilmiş matris Tablo 2'deki gibi hesaplanır.

Tablo 2. Sepal length için normalize edilmiş matris (Normalized matrix for the sepal length)

		Ölçütler		
	Bileşen sayısı	**AIC**	**AICc**	**BIC**
Alternatifler	**1**	0.4519	0.4511	0.4370
	2	0.4442	0.4438	0.4401
	3	0.4457	0.4455	0.4450
	4	0.4446	0.4450	0.4509
	5	0.4495	0.4506	0.4626

Normalize edilmiş matrisinin her bir sütunundaki elemanlar ilgili W_j ağırlık vektörü ile çarpılarak ağırlıklandırılmış normalize matris elde edilir. TOPSIS yönteminin tek sübjektif aşaması burasıdır. Burada bilgi kiriterlerinin ağırlıkları eşit olarak değerlendirilmiş olup ağırlık vektörü $W = (w_1, w_2, w_3) = \{0.3333, 0.3333, 0.3333\}$ olarak alınmıştır. Böylelikle ağırlıklandırılmış normalize matris Tablo 3'deki gibi elde edilmiştir.

Tablo 3. Sepal length için ağırlıklandırılmış normalize edilmiş matris (Weighted normalized matrix for the sepal length.)

		Ölçütler		
	Bileşen sayısı	**AIC**	**AICc**	**BIC**
Alternatifler	**1**	0.1506	0.1504	0.1457
	2	0.1481	0.1479	0.1467
	3	0.1486	0.1485	0.1483
	4	0.1482	0.1483	0.1503
	5	0.1498	0.1502	0.1542

Amacımız minimizasyon olup ağırlıklandırılmış normalize matrisin her bir sütuna ait minimum değerler ideal çözümleri $A^+ = \{0.1481, 0.1479, 0.1457\}$, maksimum değerler ise negatif ideal çözümleri $A^- = \{0.1506, 0.1504, 0.1542\}$ oluşturur. Böylece her alternatifin ideal çözüme göreli yakınlığı C^+ değerleri ve alternatiflerin sıra değerleri Tablo 4'deki gibi hesaplanmıştır.

Tablo 4. Sepal length için alternatiflerin göre ideal çözüme göreli yakınlık değerleri (Relative proximity values of alternatives to ideal solution for the sepal length)

Alternatifler	C^+	Sıralama
1	0.7063	2
2	**0.8894**	1
3	0.7007	3
4	0.5191	4
5	0.0838	5

Tablo 4'e göre en yüksek C^+ değerine (0.8894) sahip olan 2, alternatifler içerisinde en uygun bileşen sayısı olarak belirlenmiştir. Böylece Iris veri setinin 1. değişkenindeki (Sepal length) normal karma modeldeki uygun bileşen sayısı TOPSIS yöntemine göre 2 olarak bulunmuştur.

Benzer şekilde İris veri setinin ikinci, üçüncü ve dördündü değişkenlere göre elde karar matrisi ve TOPSIS yöntemine ile hesaplanan her alternatifin ideal çözüme göreli yakınlığı (C^+)**,** Tablo 5-7'de verilmiştir.

Tablo 5. Sepal width için bileşen sayısına göre bilgi kriterleri değerlerinin oluşturduğu karar matrisi ve C^+ değerleri (Decision matrix and C^+ values created by the information criteria values for the sepal width according to the number of components)

Bileşen sayısı	AIC	AICc	BIC	C^+
1	179.55	179.63	185.57	**1.0000**
2	180.38	180.66	192.43	0.8138
3	183.08	183.67	201.14	0.5560
4	187.12	188.14	211.21	0.2372
5	189.18	190.76	219.28	0.0000

Tablo 5'e göre en yüksek C^+ değerine (1.000) sahip olan 1, alternatifler içerisinde en uygun bileşen sayısı olarak belirlenmiştir. Böylece Iris veri setinin 2. değişkenindeki (Sepal width) normal karma modeldeki uygun bileşen sayısı TOPSIS yöntemine göre 1 olarak bulunmuştur.

Tablo 6. Petal length için bileşen sayısına göre bilgi kriterleri değerlerinin oluşturduğu karar matrisi ve C^+ değerleri (Decision matrix and C^+ values created by the information criteria values for the petal length according to the number of components)

Bileşen sayısı	AIC	AICc	BIC	C^+
1	599.17	599.26	605.20	0.0000
2	411.16	411.57	426.21	**1.0000**
3	415.74	416.76	439.83	0.9537
4	421.66	423.57	454.77	0.9012
5	427.50	430.61	469.65	0.8494

Tablo 6'ya göre en yüksek C^+ değerine (1.000) sahip olan 2, alternatifler içerisinde en uygun bileşen sayısı olarak belirlenmiştir. Böylece Iris veri setinin 3. değişkenindeki (Petal length) normal karma modeldeki uygun bileşen sayısı TOPSIS yöntemine göre 2 olarak bulunmuştur.

Tablo 7. Petal width için bileşen sayısına göre bilgi kriterleri değerlerinin oluşturduğu karar matrisi ve C^+ değerleri (Decision matrix and C^+ values created by the information criteria values for the petal width according to the number of components)

Bileşen sayısı	AIC	AICc	BIC	C^+
1	347.23	347.31	353.25	0.0000
2	220.81	221.23	235.86	**0.9810**
3	217.66	218.68	241.74	0.9746
4	228.72	229.74	252.81	0.8959
5	232.72	234.30	262.83	0.8453

Tablo 7'ye göre en yüksek C^+ değerine (0.9810) sahip olan 2, alternatifler içerisinde en uygun bileşen sayısı olarak belirlenmiştir. Böylece Iris veri setinin 4. değişkenindeki (Petal width) normal karma modeldeki uygun bileşen sayısı TOPSIS yöntemine göre 2 olarak bulunmuştur.

TOPSIS yöntemi ile değişkenlerin bileşen sayıları belirlenmiş 2. değişkende (Sepal width) bölünme olmadığından (tek bileşene sahip) kümelemeye dâhil edilmemiştir. Böylece boyut indirgenmiş veri setinde bileşen sayısı belirlenirken 3 değişken (Sepal length, petal length ve petal width) ve bu değişkenlerdeki bileşen sayılarına göre hesaplanmıştır. Değişkenlerdeki bileşen sayılarına göre karma modellerde oluşabilecek minimum $k_{min} = max\{k_s\} = \{2,1,2,2\} = 2$ ve maksimum küme sayısı $k_{max} = \prod_{s=1}^{p} k_s = 2x1x2x2 = 8$ olacak şekilde elde edilir. Karma modeldeki küme sayısı $2 \leq k \leq 8$ aralığında kalmak üzere belirlenmiştir.

İndirgenmiş Iris veri setinde 1., 3. ve 4. değişkenlere GMM uygulayarak elde edilen bileşen sayısına göre bilgi kriterleri değerlerinin oluşturduğu karar matrisi ve TOPSIS yöntemine göre hesaplanan C^+ değerleri**,** Tablo 8'de verilmiştir.

Tablo 8. İndirgenmiş Iris veri setinde için alternatif bileşen sayısına göre bilgi kriterleri değerlerinin oluşturduğu karar matrisi ve C^+ değerleri (Decision matrix and C^+ values formed by the values of information criteria according to the number of alternative components for the reduced Iris data set)

Bileşen Sayısı	AIC	AICc	BIC	C^+
2	457.07	461.71	508.26	0.3971
3	423.86	434.84	**499.12**	**0.8731**
4	**415.37**	**434.71**	514.72	0.8150
5	435.50	450.00	522.81	0.5027
6	430.02	450.72	532.39	0.4786
7	430.46	458.82	547.87	0.3547
8	415.87	453.58	548.33	0.4863

Tablo 8'e göre, AIC, AICc ve BIC bilgi kriterlerinin sırasıyla indirgenmiş veri için küme sayısı tahminleri 4, 4 ve 3 olarak bulunmuştur. Önerilen yöntem ile uygun küme sayısı ise en yüksek C^+ değerine (0.8731) sahip olan 3, alternatifler içerisinde en uygun bileşen sayısı olarak belirlenmiştir. Böylece indirgenmiş Iris veri setinin ideal küme sayısı olacak şekilde üç kümeli model bulunmuştur. Dört değişkenli Iris veri setinde mevcut küme sayısı üç olduğundan önerilen yöntem küme sayısını doğru belirlemiştir.

Modelin uygunluğu, üyelikleri bilinen gözlemlerin doğru kümeye atanma oranları olan DSO üzerinden karşılaştırma yapılarak ölçülmüştür. İris veri setindeki tüm değişkenler göz önünde bulundurulan tam veri seti ve üç değişkene düşürülmüş indirgemiş veri seti GMM ile 3 bileşenli olarak modellendiğinde elde edilen atama sayıları Tablo 9'da verilmiştir.

Tablo 9. Iris veri setinde tam ve indirgenmiş verideki gözlemlerin çapraz tablo değerleri (Cross tabular values of observations in complete and reduced data in Iris dataset)

			Gerçek Üyelikler		
			setosa	versicolor	virginica
Model Atama Üyelikleri	**Tam Veri**	setosa	50	0	0
		versicolor	0	50	5
		virginica	0	0	45
	İndirgenmiş Veri	setosa	50	0	0
		versicolor	0	50	3
		virginica	0	0	47

Tablo 9'a göre tam ve indirgenmiş veri setinde setosa ve versicolor kümeleri için tam atama yapılırken, tam veride virginica kümesine ait 5 gözlem, indirgenmiş veride de 3 gözlem versicolor kümesine yanlış olarak atanmıştır. Tam ve indirgenmiş veri seti için hesaplanan DSO'lar Tablo 10'da verilmiştir

Tablo 10. Iris veri setinde tam ve indirgenmiş veriye göre %DSO'lar (%DSOs according to complete and reduced data in Iris data set)

		setosa	**versicolor**	**virginica**
Tam Veri	**DSO**	%100	%100	%90
	Genel DSO		%96.7	
İndirgenmiş Veri	**DSO**	%100	%100	%94
	Genel DSO		%98	

Tablo 10'daki tam ve indirgenmiş veride hesaplanan DSO lara göre TOPSIS ile belirlenen bileşenler ve boyut indirgeme sonucu daha yüksek başarı oranı yani daha iyi model uygunluğu elde edilmiştir.

5. Sonuç ve Tartışma (Result and Discussion)

Çok değişkenli verilerin küme sayısının tahmini için kümeleme analizinde birçok yöntem kullanılmaktadır. Kullanılan yöntemler küme sayısının alt ve üst sınırı için etkili bir kriter belirleyememiştir. Kümeleme analizinde küme sayısını doğru tahmin etmek için yaygın olarak bilgi kriterlerinden yararlanılmaktadır. Bu çalışmada çok değişkenli veride oluşabilecek alternatif küme sayısı için bir aralık önerilmiş ve küme sayısını bilgi kriterlerinin yanı sıra deterministik yaklaşımlar ile belirleme amaçlanmıştır. Değişkenlerdeki bileşen sayılarına dayalı küme sayısı tahmini için, her bileşenin AIC, AICc ve BIC bilgi kriterleri ön bilgi olarak verilmiş, TOPSIS yöntemi bileşen sayıları bilgi kriterlerine göre daha doğru elde edilmiştir. Bileşen sayılarından küme sayılarını modele dayalı elde etmek için bir aralık bulunmuş ve normal karma modellerle daha yüksek başarı oran ile kümelenmiştir. Önerilen yöntem ile veri setinde bileşen bulunmayan değişkenler elenip boyut indirgenmiş böylece tam veriye göre indirgenmiş verideki kümeleme başarısı artmıştır. Kümeleme literatüründe yaygın olarak kullanılan İris veri seti üzerinde önerilen yöntem ile boyut indirgeyip DSO %98 olarak elde edilmiştir. Böylece uygulanan yöntemdeki boyut indirgeme ve değişken seçiminin sınıflandırma başarısı üzerine etkisi ortaya konulmuştur.

Bu çalışma, çok değişkenli dağılımların karmasına dayalı kümeleme analizinde uygun küme sayısının seçimi ve kümeleme için en uygun değişken kümesinin belirlenmesi ile ilgili yeni bir yaklaşım ortaya koymuştur. Öncelikle veri setinde yer alan her bir değişken için uygun bileşen sayısını, tek değişkenli normal dağılımların karmasına dayalı kümelemede bilgi kriterlerine dayalı olarak TOPSIS yöntemi ile belirlenmiştir. Boyut indirgeme işlemi için bileşen sayısı 1 olarak belirlenen değişkenleri dışlayarak gerçekleştirilmiştir. Daha sonra uygun küme sayısı için hesaplanan aralıkta bilgi kriterlerine dayalı olarak TOPSIS yöntemi ile uygun küme sayısı için nokta tahmini belirlenmiştir. Uygun küme sayısı belirlendikten sonra, indirgenmiş veride optimum kümeleme ortaya konulmuştur.

Çıkar Çatışması (Conflict of Interest)

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir. No conflict of interest was declared by the authors.

Kaynaklar (References)

Acer, A., Kalender, S., 2020. Antrepoların Performansının Entropi ve TOPSIS Yöntemiyle Değerlendirilmesi. Sosyal Bilimler Dergisi/Journal of Social Sciences, (65).

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6): 716–723.

Akalp, G., Özok, A., 2017. Ergonomik Risklerin Bulanik Mantik Yöntemi Ile Modellenmesi Ve Bir Uygulama, Mühendislik Bilimleri Ve Tasarim Dergisi, 5 (0), 69-79

Akogul, S., 2018. Çok Değişkenli Verilerin Modele Dayalı Kümeleme Analizinde Kümeleme Etkinliğinin Arttırılması", Doktora Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.

Akogul, S., Erisoglu, M., 2016. A comparison of information criteria in clustering based on mixture of multivariate normal distributions. Mathematical and Computational Applications, 21(3), 34.

Akogul, S., Erisoglu, M., 2017. An approach for determining the number of clusters in a model-based cluster analysis. Entropy, 19(9), 452.

Andriyanov, N., Tashlinsky, A., Dementiev, V., 2020. Detailed Clustering Based on Gaussian Mixture Models. In Proceedings of SAI Intelligent Systems Conference (pp. 437-448). Springer, Cham.

Binder, D. A., 1978. Bayesian cluster analysis. Biometrika, 65(1), 31-38.

Bozdogan, H., 1994. Mixture-model cluster analysis using model selection criteria and a new informational measure of complexity, Proceedings of the first US/Japan conference on the frontiers of statistical modeling: An informational approach, 69-113.

Burak, E., Boran, F., Mustafa, K., 2015. Sezgisel Bulanik TOPSIS Yöntemi Kullanilarak Ergonomik Ürün Konsept Seçimi. Mühendislik Bilimleri ve Tasarım Dergisi, 3(3), 433-440.

Çetin, M. H., Alvalı, G. T., 2020. Yük Vagonu Bojisi Tasarımında Çok Kriterli Karar Verme Teknikleri İle Malzeme Seçimi, Mühendislik Bilimleri ve Tasarım Dergisi, 8(1), 91-104.

Day, N. E., 1969. Estimating the components of a mixture of normal distributions. Biometrika, 56(3), 463-474.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39(1), 1-22.

Ece, N., 2019. Holding Şirketlerinin Finansal Performans Sıralamasının Entropi Tabanlı TOPSIS Yöntemleri İle İncelenmesi. Finans Ekonomi ve Sosyal Araştırmalar Dergisi (FESA), 4(1), 63-73.

Fisher, R. A., 1936. The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics, 7(2): 179-188.

Fop, M., Murphy, T. B., 2018. Variable selection methods for model-based clustering. Statistics Surveys, 12, 18-65.

Fraley, C. and Raftery, A. E., 1998. How Many Clusters? Which Clustering Method? Answers via Model-Based Cluster Analysis. The Computer Journal, 41, 578-588.

Fraley, C., Raftery, A. E., 2002. Model-based clustering, discriminant analysis, and density estimation, Journal of the American statistical Association, 97 (458), 611-631.

Galimberti, G., Soffritti, G., 2007. Model-based methods to identify multiple cluster structures in a data set. Computational statistics & data analysis, 52(1), 520-536.

Gogebakan, M., 2017. Karma Dağılım Modelleri Kullanılarak Çok Değişkenli Veride Grup Yapılarının Belirlenmesi, Ayrıştırılması, Kümelenmesi ve Sınıflandırılması, Doktora Tezi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Kayseri.

Gogebakan, M., Erol, H., 2019. Mixture Model Clustering Using Variable Data Segmentation and Model Selection: A Case Study of Genetic Algorithm, Mathematics Letters. Vol. 5, No. 2, 2019, pp. 23-32

Gogebakan, M., Erol, H., 2019. Normal Mixture Model-Based Clustering of Data Using Genetic Algorithm. In The International Conference on Artificial Intelligence and Applied Mathematics in Engineering (pp. 539-543). Springer, Cham.

Gögebakan, M., Servi, T., 2019. Genetik Algoritma Kullanılarak Verilerin Karma Normal Modele Dayalı Kümelenmesi. Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi, 35(3), 12-23.

Hurvich, C. M., Tsai, C. L., 1989. Regression and time series model selection in small samples. Biometrika, 76(2), 297-307.

Hwang, C. L., Yoon, K., 1981. Methods for multiple attribute decision making. In Multiple attribute decision making (pp. 58-191). Springer, Berlin, Heidelberg.

Ishizaka, A., Nemery, P., 2013. Multi-criteria decision analysis: methods and software, John Wiley & Sons, p.

Khan, B. M., Bilal, R., Young, R., 2018. Fuzzy-TOPSIS based cluster head selection in mobile wireless sensor networks. Journal of Electrical Systems and Information Technology, 5(3), 928-943.

McLachlan, G. J., Chang, S. U., 2004. Mixture Modelling for Cluster Analysis. Statistical Methods in Medical Research 13, 347-361.

McLachlan, G., Peel, D., 2004. Finite mixture models, John Wiley & Sons.

Mirzal, A., 2020. Statistical Analysis of Microarray Data Clustering using NMF, Spectral Clustering, Kmeans, and GMM. IEEE/ACM Transactions on Computational Biology and Bioinformatics.

Nguyen, H. D., McLachlan, G. J., 2015. Maximum likelihood estimation of Gaussian mixture models without matrix operations. Advances in Data Analysis and Classification, 9(4), 371-394.

Oliveira-Brochado, A., Martins, F. V., 2005. Assessing the number of components in mixture models: a review, Universidade do Porto, Faculdade de Economia do Porto.

Özgüner, Z., 2020. Dış Kaynak Kullanımı Kapsamında Entegre Entropi-TOPSIS Yöntemleri ile Tedarikçi Seçimi Probleminin Çözümlenmesi. İşletme Araştırmaları Dergisi, 12(2), 1109-1120.

Pearson, K., 1894. Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London. A, 185, 71-110.

Perçin, S., Sönmez, Ö., 2018. Bütünleşik Entropi Ağirlik Ve TOPSIS Yöntemleri Kullanilarak Türk Sigorta Şirketlerinin Performansinin Ölçülmesi. Uluslararası İktisadi ve İdari İncelemeler Dergisi, (18. EYİ Özel Sayısı), 565-582.

Ridolfi, A., Idier, J., 2001. Penalized maximum likelihood estimation for univariate normal mixture distributions. In AIP Conference Proceedings (Vol. 568, No. 1, pp. 229-237). American Institute of Physics.

Schwarz, G., 1978. Estimating the dimension of a model, Ann. Statist. 6 pp. 461–464.

Seo, B., Kim, D., 2012. Root selection in normal mixture models. Computational Statistics & Data Analysis, 56(8), 2454-2470.

Servi, T., 2009. Çok değişkenli karma dağılım modeline dayalı kümeleme analizi, Doktora Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana

Soffritti, G., 2003. Identifying multiple cluster structures in a data matrix. Communications in Statistics-Simulation and Computation, 32(4), 1151-1177.

Şahin, Y., Supçiller, A., 2015. Tedarikçi seçimi için bir karar destek sistemi. Mühendislik Bilimleri ve Tasarım Dergisi, 3(2), 91-104.

Wang, T. C., Lee, H. D., 2009. Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. Expert systems with applications, 36(5), 8980-8985.

Yıldırım, B. F., Önder, E., 2015. Çok kriterli karar verme yöntemleri. Bursa: Dora Basım-Yayın Dağıtım.