

# KÜMELEME ÇÖZÜMLEMESİNDE DÜZELTİLMİŞ TEK ADIM M-TAHMİN EDİCİSİNİN KULLANILMASI

Abdullah Fırat ÖZDEMİR\*

Engin YILDIZTEPE\*\*

## ÖZET

*Kümeleme çözümlemesi, yaygın olarak kullanılan çok değişkenli istatistiksel yöntemlerden biridir. Düzeltilmiş tek adım M-tahmin edicisi ise Huber'in (Huber, 1964) önerdiği M-tahmin edicisi üzerinde Wilcox tarafından yapılan değişiklik ile elde edilmiştir (Wilcox, 2003a). Düzeltilmiş tek adım M-tahmin edicisi, 2003 yılından bugüne kadar varyans çözümlemesi (Wilcox ve Keselman, 2003) ve çoklu karşılaştırmalar (Wilcox, 2003b) gibi alanlarda kullanılmıştır. Hiyerarşik olmayan kümeleme yöntemi olarak sınıflandırılan k-ortalama yöntemini konu alan bu çalışmada, ortalama yerine düzeltilmiş tek adım M-tahmin edicisi (Modified one-step M-estimator-MoM) kullanılarak geliştirilen kümeleme algoritması tanıtılmıştır.*

**Anahtar Kelimeler:** Düzeltilmiş tek-adım M-tahmin edicisi, k-Ortalama kümeleme, Kümeleme çözümlemesi.

## 1. GİRİŞ

Kümeleme çözümlemesinin genel amacı, gruplanmamış verileri benzerliklerine göre sınıflandırmak (gruplamak) ve araştırmacıya uygun, işe yarar özetleyici bilgiler elde etmede yardımcı olmaktır (Tatlídil, 2002, 329). Kümeleme çözümlemesinde küme sayısı önceden bilinmemekte ve sadece verilerin mevcut durumuna ilişkin sonuçlar alınmaktadır. Benzerlik ölçüsü olarak veriler arasında hesaplanan uzaklık değerleri kullanılmaktadır. Literatürde Minkowski, Mahalanobis, Canberra, Hotelling  $T^2$  ve Öklit gibi uzaklık türleri kullanılmakla beraber bu çalışmada Öklit uzaklığı kullanılmıştır (Dallas, 1998; Han ve Kamber, 2001; Tatlídil, 2002). Kümeleme çözümlemesi çoğunlukla resim tanıma, veri madenciliği, metin madenciliği, uzaysal veritabanı uygulamaları, müşteri ilişkileri yönetimi, sağlık ve biyoloji gibi alanlarda kullanılmaktadır.

Kümeleme çözümlemesinde kullanılan kümeleme yöntemleri, hiyerarşik olan ve hiyerarşik olmayan yöntemler olarak ikiye ayrılır. Küme sayısı konusunda bir ön bilgi var ise veya araştırmacı anlamlı olacak küme sayısına karar vermiş ise kuramsal dayanakları daha güçlü olan hiyerarşik olmayan yöntemler tercih edilmektedir (Tatlídil, 2002, 338). Bu çalışmada konu edilen k-ortalama yöntemi en çok tercih edilen hiyerarşik olmayan yöntemlerden biridir.

\* Yrd. Doç. Dr. A. Fırat Özdemir, Dokuz Eylül Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü, e-posta: [firat.ozdemir@deu.edu.tr](mailto:firat.ozdemir@deu.edu.tr)

\*\* Öğr. Gör. Dr. Engin Yıldıztepe, Dokuz Eylül Üniversitesi Fen-Edebiyat Fakültesi Bilgisayar Bilimleri Bölümü, e-posta: [engin.yildiztepe@deu.edu.tr](mailto:engin.yildiztepe@deu.edu.tr)

Çalışmanın 2. bölümünde k-ortalama algoritması, düzeltilmiş tek adım M-tahmin edicisi ve k-MoM algoritması tanıtılmıştır. 3. bölümde, önerilen algoritma için geliştirilen yazılım ve kullanılan veri setleri tanıtılmış, 4. bölümde ise bu veri setleri ile yapılan uygulamaların sonuçları verilmiştir. Son bölüm, elde edilen sonuçlar ile yapılan tartışmaları içermektedir.

## 2. YÖNTEM

### 2.1 K-Ortalama Algoritması, Düzeltilmiş Tek Adım M-Tahmin Edicisi ve K-MoM Algoritması

#### 2.1.1 k-Ortalama Algoritması

Algoritma 1967 yılında J.B. MacQueen tarafından geliştirilmiştir (MacQueen, 1967). En çok kullanılan algoritmalarından birisidir ve her veri sadece bir kümeye dâhil olabilir. Eşit büyüklükte küresel kümeler bulma eğiliminde olan algoritma 4 adımda uygulanır;

1. İlk küme merkezleri olarak birimler arasından rastgele küme sayısı kadar (k) başlangıç noktası belirlenir.
2. Her birimin seçilen merkez noktalara olan uzaklıkları hesaplanır. Sonuçlara göre tüm birimler k adet kümeden kendilerine en yakın olan kümeye atanır.
3. Oluşan kümelerin yeni merkez noktaları o kümedeki tüm birimlerin aritmetik ortalamaları ile değiştirilir.
4. Merkez noktalar değişmeye kadar 2. ve 3. adımlar tekrarlanır.

k-ortalama algoritmasının daha çok büyük veri setlerinde etkili sonuçlar verdiği bilinmektedir (Han ve Kamber, 2001). Algoritmanın en büyük dezavantajlarından birisi aykırı değerlerden çok etkilenmesidir (Richard ve Dean, 1992; Fasulo, 1999; Han ve Kamber, 2001). Literatürde, kümelemenin daha güvenilir yapılabilmesi için veri seti içinde yer alan aykırı değerlerin, oluşan kümelerin dışında tutulabileceğini belirten çalışmalar görülmektedir (Guha vd., 1998; Hautamäki vd., 2005). Bu çalışmada k-ortalama algoritmasında üçüncü adımda yer alan aritmetik ortalama yerine Wilcox tarafından önerilen (Wilcox, 2003a) düzeltilmiş tek adım M-tahmin edicisi kullanılmış ve sonuçları tartışılmıştır.

#### 2.1.2 Düzeltilmiş Tek Adım M-Tahmin Edicisi

Bir serinin merkezi, o serideki gözlem değerlerinin serinin merkezine olan uzaklıklarını en küçükleyen yerdedir.  $X_1, \dots, X_n$  n birimlik rassal örneklem,

$\psi(x)$ : seride yer alan değerlerin merkeze olan uzaklığını ölçen fonksiyon ve  
c: konum ölçüsü (merkez nokta) olmak üzere,

$$\psi(X_1 - c) + \psi(X_2 - c) + \dots + \psi(X_n - c) = 0 \quad (1)$$

eşitliği kullanılarak türetilen konum ölçüleri M-tipi konum ölçüsü ya da kısaca seri merkezinin M-tahmin edicisi olarak adlandırılır (Huber, 1964). Burada  $\psi(x) = x$  fonksiyonu kullanılırsa aranan c konum ölçüsü,

$$\begin{aligned} (X_1 - c) + (X_2 - c) + \dots + (X_n - c) &= 0 \\ X_1 + X_2 + \dots + X_n &= nc \\ \frac{X_1 + X_2 + \dots + X_n}{n} &= c \end{aligned} \quad (2)$$

biçiminde görüldüğü gibi aritmetik ortalamaya eşit olur. Huber'in önerdiği  $\psi(x)$  fonksiyonu ile M-tahmin edicisi Newton-Raphson gibi adimsal yöntemler ve bilgisayar yardımı ile hesaplanabilir. Yöntemin bir adım çalıştırılması ile elde edilen tek-adım M-tahmin edicisi

$$\hat{\mu}_{os} = \frac{K(MADN) + (i_2 - i_1) \sum_{i=i_1+1}^{n-i_2} X_{(i)}}{n - i_1 - i_2} \quad (3)$$

biçimindedir. Burada  $i_1$  ve  $i_2$  aşağıda eşitlik 5 ve 6'da verildiği gibidir ve K değeri 1.28'e eşittir. Wilcox, küçük örneklem genişliği ile çalışılması durumunda uygulamada karşılaşılan problemleri gidermek amacı ile yaptığı bir değişiklik ile düzeltilmiş tek-adım M-tahmin edicisini önermiştir (Wilcox, 2003a).

$X_{(i)}, X_i$  serisinin küçükten büyüğe sıralanmış biçimi olmak üzere düzeltilmiş tek-adım M-tahmin edicisi,

$$\hat{\mu}_{mom} = \frac{\sum_{i=i_1+1}^{n-i_2} X_{(i)}}{n - i_1 - i_2} \quad (4)$$

eşitliği ile hesaplanır. Burada,

$$i_1 : \left( \frac{(X_i - M)}{MADN} \right) < -2.24 \quad (5)$$

şartını sağlayan gözlem sayısı ve

$$i_2 : \left( \frac{(X_i - M)}{MADN} \right) > 2.24 \quad (6)$$

şartını sağlayan gözlem sayısıdır. Burada M medyan olmak üzere,

$$MADN = \frac{MAD}{0,6745} \quad (7)$$

ve

$$MAD = med(|X_1 - M|, \dots, |X_n - M|) \quad (8)$$

olarak tanımlanır.

### 2.1.3 k-Düzeltilmiş Tek-Adım M-Tahmin Edicisi Algoritması

Çalışmada önerilen k – düzeltilmiş tek-adım M-tahmin edicisi (k-MoM) algoritması beş adımdan oluşmaktadır. k-ortalama algoritmasından farklı olarak beşinci adımda

$$\frac{|(X_i - M)|}{MADN} > 2.24 \quad (9)$$

biçiminde tanımlanan ve Hampel belirteci olarak bilinen bir aykırı değer belirleme kuralı kullanılmıştır. Algoritma, ilk iki adımı 2.1.1 nolu alt başlıkta verilen algoritma ile aynı olmak üzere, 3. adımdan itibaren aşağıda verilmiştir.

3. Oluşan kümelerin yeni merkez noktaları o kümedeki tüm birimlerin düzeltilmiş tek adım M-tahmin edicisi ile değiştirilir.
4. Merkez noktalar değişmeyene kadar 2. ve 3. adımlar tekrarlanır.
5. Her bir küme için küme içi kareler toplamı hesaplanır.  $|X_i - M| / MADN > 2.24$  koşulu gerçekleşen gözlem değerleri bu hesaplama katılmaz ve çizilen küme grafiğinde bu kümeye ait bir gözlem olarak görünmez.

## 2.2 Kullanılan Bilgisayar Yazılımı ve Veri Setleri

Önerilen k-MoM kümeleme algoritması için geliştirilen bilgisayar yazılımında k-ortalama algoritması da çalıştırılabilmektedir. k-MoM yazılımı Delphi 7.0 uygulama geliştirme aracı ile geliştirilmiştir.

### 2.2.1 Veri Kaynağının Seçilmesi

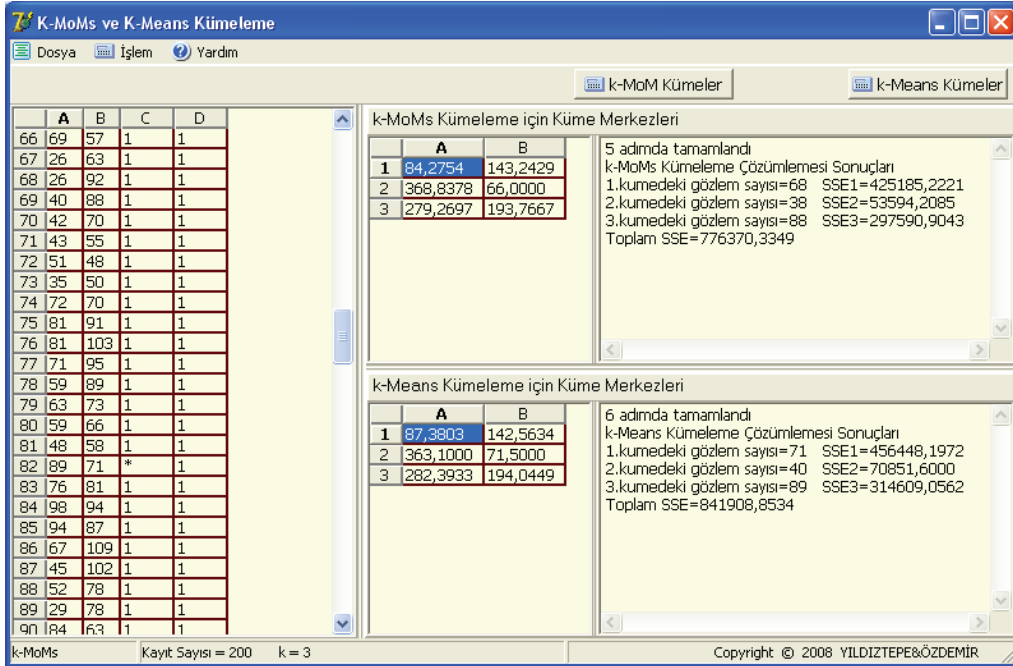
Yazılım çalıştırıldığında kullanıcının karşısına ilk olarak Şekil 1’de görülen pencere gelir. Kullanıcı bu pencereyi kullanarak öngördüğü küme sayısını (k) ve veri kaynağını belirler. Küme sayısı alanına ancak tamsayı değerler girilebilir ve birden küçük değer girilemez. Yazılımda, değişken ve kayıt sayıları belirtilerek veri girişi yapılabileceği gibi metin dosyalarında bulunan veriler de kullanılabilir. Eğer veriler bir dosyada kayıtlı ise “Dosyadan” seçeneği seçilerek “Tamam” butonuna basılır, böylece verileri içeren csv uzantılı (veriler virgülle ayrılmış) dosya kullanılabilir. Değişken ve kayıt sayılarını belirterek veri girişi yapmak için “Veri Girişi” seçeneği işaretlenmelidir. Daha sonra aktif hale gelen “Değişken Sayısı” ve “Kayıt Sayısı” alanlarına veri kümesindeki değişken sayısının ve kayıt sayısının yazılması gereklidir. “Veri Girişi” seçeneği işaretlendiğinde; değişken sayısı ve kayıt sayısı alanları boş geçilemez, değişken sayısı olarak birden, kayıt sayısı olarak ikiden küçük bir sayı girilemez. Bu alanlarda belirtilen değişken ve kayıt sayılarına göre, veriler açılan ana form üzerindeki tabloya girilebilir.

Şekil 1. Veri Kaynağı Seçimi

### 2.2.2 Yazılımın Kullanılması

Veri kaynağı seçildikten sonra Şekil 2’de verilen ana formda kümeleme algoritması çalıştırılabilir. Eğer veri kaynağı olarak bir dosya seçilmişse, veriler formun sol tarafındaki tabloya aktarılır. Başlangıç kümeleri bu veriler arasından belirtilen küme sayısı kadar rastgele seçilir ve k-MoM ve k-ortalama algoritmalarında kullanılmak üzere sağ tarafta bulunan tablolara aktarılır. Veri kaynağı seçimi formunda veri girişi seçilmişse, sol taraftaki tabloya veriler ve sağ taraftaki tablolara başlangıç kümeleri girilmelidir.

Veri girişi tamamlandıktan veya istenilen veri dosyası seçildikten sonra kullanıcı algoritmaları çalıştırabilir. k-MoM kümeler butonu k-MoM algoritmasını, k-Means kümeler butonu k-ortalama algoritmasını çalıştırır. Algoritma sonucunda bulunan kümelerdeki eleman sayıları, küme içi değişim değerleri ve bunların toplamı son küme merkezi değerlerini gösteren tabloların sağında belirtilir. Formun sol tarafında verilerin bulunduğu tabloda başlangıçta iki boş sütun daha vardır. Bunlardan birincisi k-MoM kümeleme sonucu, ikincisi k-ortalama kümeleme sonucu o gözlemin hangi kümeye ait olduğunu gösterir. k-MoM kümeleme sonucunda gözlemin ait olduğu kümeyi belirten sütündeki ‘\*’ işareti o gözlemin aykırı değer olarak işaretlendiğini belirtir.



Şekil 2. k-MoM ve k-ortalama Kümeleme Yazılımı Ana Form

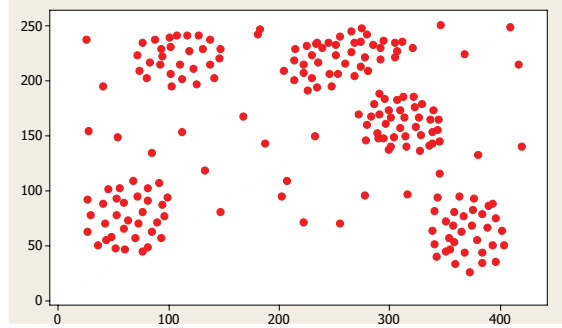
Şekil 3'te gösterilen dosya menüsü kullanılarak yeni bir veri kaynağı seçilebilir. Kaydet menüsü ile veri dosyası csv (veriler virgülle ayrılmış), algoritma sonuçları ise metin dosyası olarak kaydedilebilir. Algoritmaları çalıştırma komutları işlem menüsünden de verilebilir.



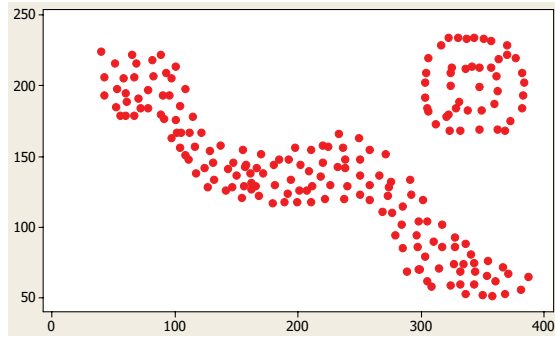
Şekil 3. Dosya Menüsü Seçenekleri

### 2.2.3 Kullanılan Veri Setleri

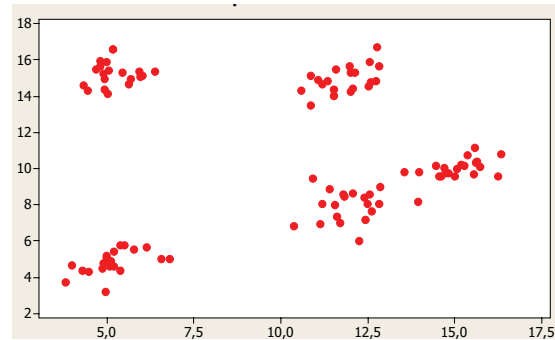
Çalışmada dört adet veri seti kullanılmıştır. Document\_Sim (Şekil 4,  $k=5$ ) ve Image Extraction (Şekil 5,  $k=2$ ) isimli sentetik veri setleri Ösmar Zaiane ve Yalling Pei tarafından hazırlanmış veri tabanlarından alınmıştır (Zaiane ve Pei, 2008). Sentetik I (Şekil 6,  $k=5$ ) ve Sentetik II (Şekil 7,  $k=3$ ) ismi verilen diğer iki veri seti ise tarafımızdan MINITAB yazılımının 14. sürümü kullanılarak üretilmiştir. Bu veri setlerinden image extraction dışındakiler aykırı değer içeren veri setleridir. Sentetik I, kümelerde az sayıda gözlem olacak şekilde, Sentetik II ise kümelerde farklı sayıda gözlem olacak şekilde üretilmiştir.



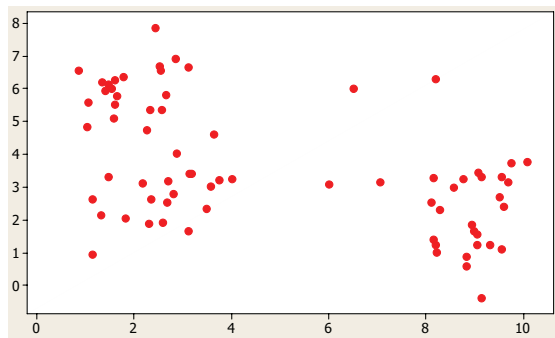
Şekil 4. Document\_Sim Veri Seti



Şekil 5. Image Extraction Veri Seti



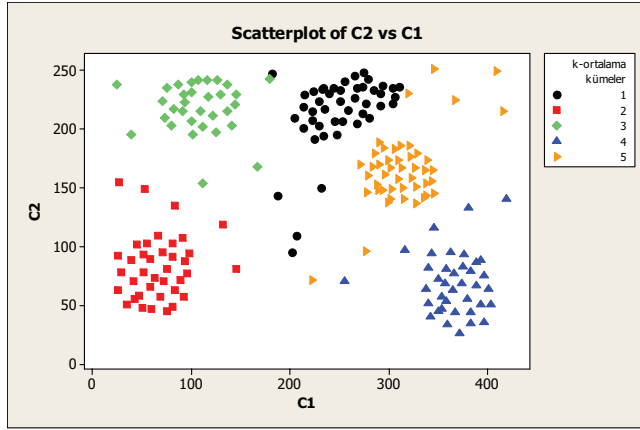
Şekil 6. Sentetik I Veri Seti



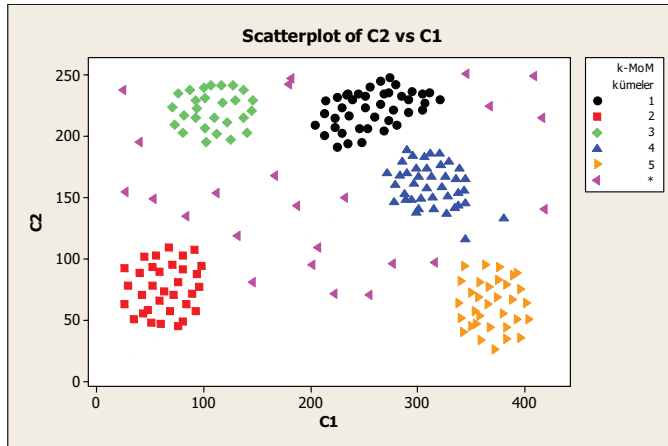
Şekil 7. Sentetik II Veri Seti

### 3. BULGULAR

Bu bölümde yapılan uygulamaların amacı; gerçekte kaç kümeye sahip olduğu bilinen veri setleri kullanarak hem önerilen algoritmanın bu kümeleri tespit etmedeki başarısının gözlenmesi, hem de alınan sonuçların k-ortalama algoritmasından alınan sonuçlar ile karşılaştırılmasıdır. Her iki algoritma için Şekil 4, 5, 6 ve 7’de verilen veri setleri ile gerçekleştirilen uygulamalar ve elde edilen kümeleme sonuçları bu bölümde verilmiştir.



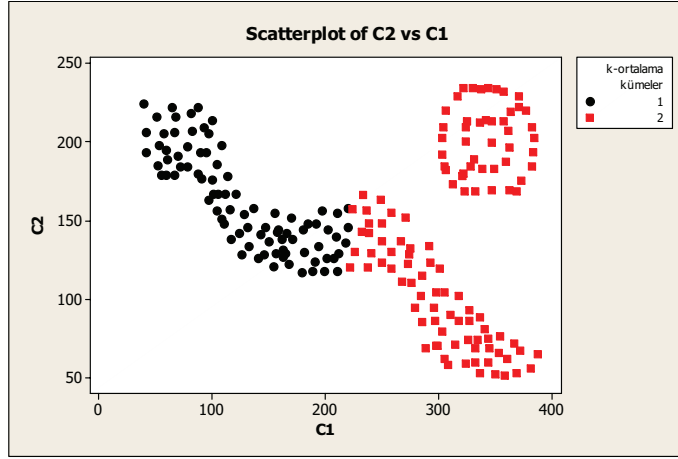
Şekil 8. Document\_Sim Veri Setinde k=5 için k-ortalama Algoritması ile Elde Edilen Sonuç



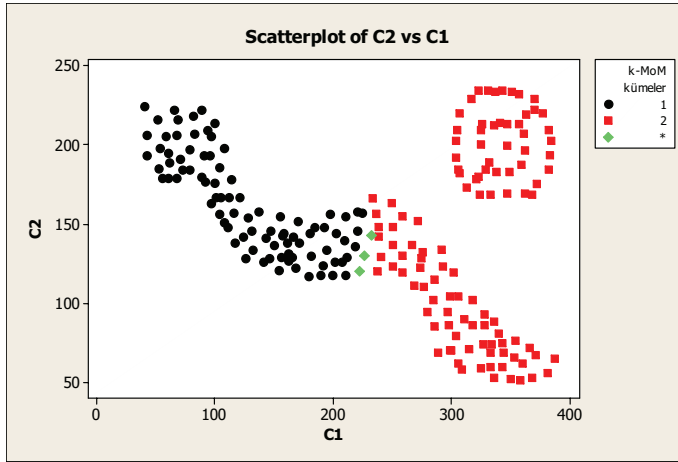
Şekil 9. Document\_Sim Veri Setinde k=5 için k-MoM Algoritması ile Elde Edilen Sonuç

K-ortalama algoritması için Şekil 8 ve k-MoM algoritması için Şekil 9 incelendiğinde document\_Sim isimli veri setinin uygulamasında k-ortalama algoritmasının aykırı değerlerden etkilendiği ve aykırı değer olduğu görülen gözlem değerlerini de kümelere dahil ettiği, k-MoM algoritmasının ise aykırı değerleri tespit etmede başarılı olduğu gözlenmektedir. Küme içi kareler toplamı (KİKT), k-ortalama için KİKT=352798,2733 değerini alırken k-MoM için KİKT=151919,4792 çıkmıştır.

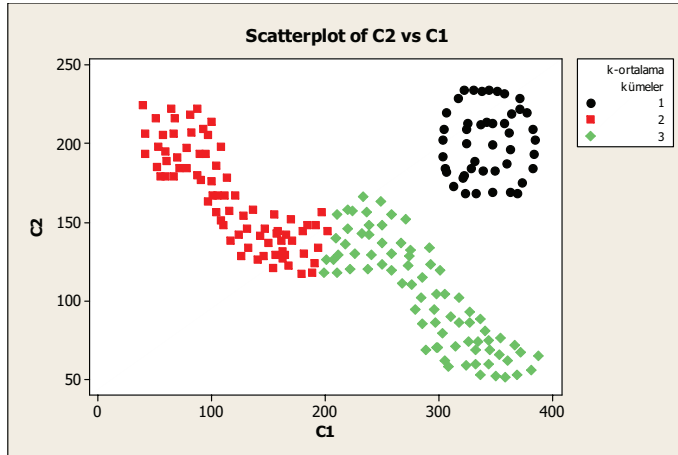




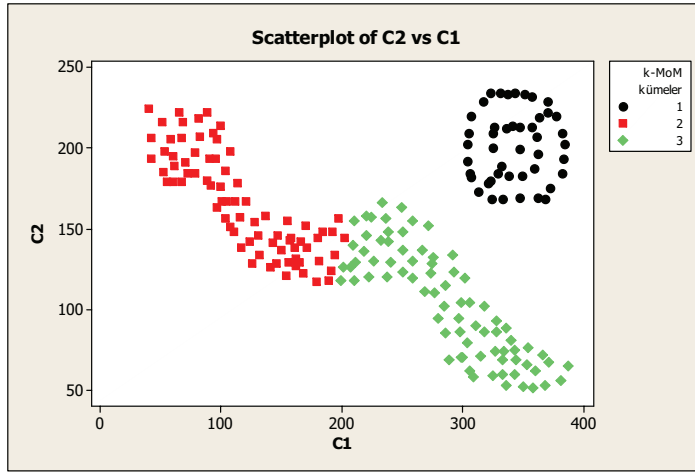
Şekil 10. Image Extraction Veri Setinde  $k=2$  için  $k$ -ortalama Algoritması ile Elde Edilen Sonuç



Şekil 11. Image Extraction Veri Setinde  $k=2$  için  $k$ -MoM Algoritması ile Elde Edilen Sonuç

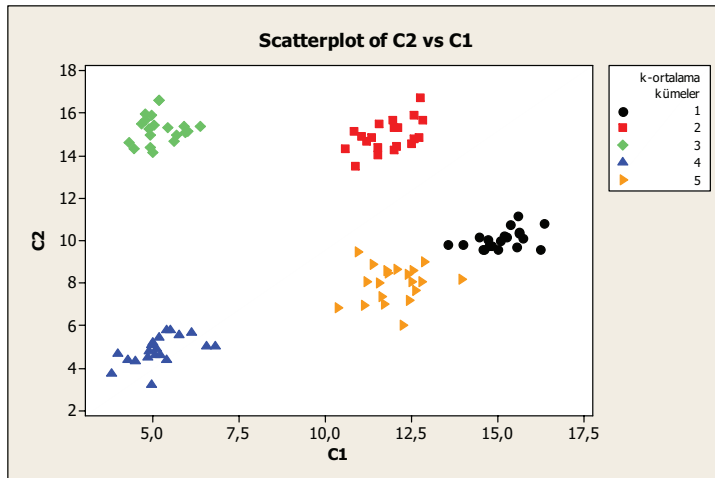


Şekil 12. Image Extraction Veri Setinde  $k=3$  için  $k$ -ortalama Algoritması ile Elde Edilen Sonuç

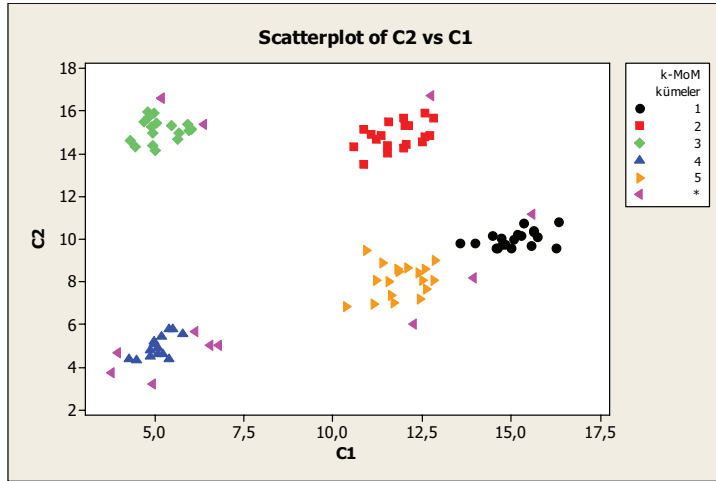


Şekil 13. Image Extraction Veri Setinde  $k=3$  için k-MoM Algoritması ile Elde Edilen Sonuç

Image Extraction veri setinin 2 kümeden oluşan bir veri seti olduğu bilinmesine rağmen Şekil 5 incelendiğinde  $k=2$  ve  $k=3$  olmak üzere 2 farklı küme sayısı için uygulama yapılmasının faydalı olacağı düşünülmüştür. Şekil 12 ve Şekil 13 incelendiğinde  $k=3$  için ortaya çıkan kümelerin aynı olduğu gözlenmiştir. Bu iki şekil için küme içi kareler toplamı değerleri de aynı çıkmıştır.  $k=2$  için Şekil 10 (k-ortalama) ve Şekil 11 (k-MoM) incelendiğinde k-MoM algoritmasının 3 gözlemi aykırı değer olarak tespit ettiği, ancak her iki algoritma tarafından belirlenen kümelerin oldukça benzer olduğu görülmüştür. Burada k-ortalama için  $KİKT=899176,5369$  ve k-MoM için  $KİKT=874216,4064$  olarak hesaplanmıştır. Elde edilen KİKT değerlerinin yakınlığı ve her iki algoritma tarafından belirlenen kümelerin benzerliği düşünüldüğünde k-MoM algoritmasının aykırı değer içermeyen veri setlerinde, k-ortalama algoritması ile benzer sonuçlar verdiği söylenebilir.

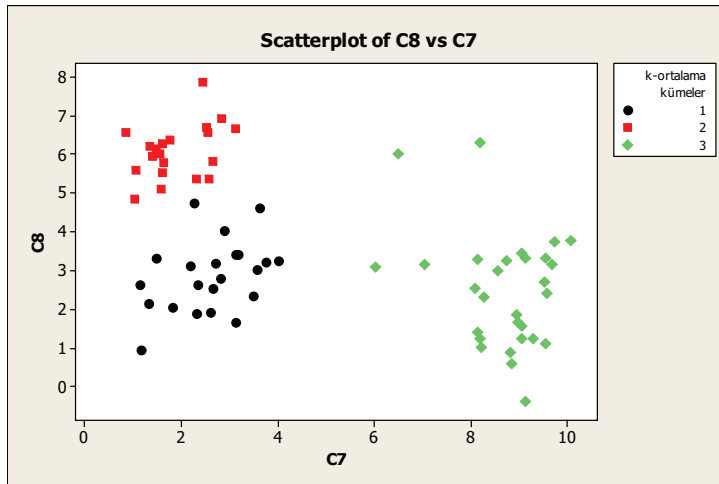


Şekil 14. Sentetik I Veri Setinde  $k=5$  için k-ortalama Algoritması ile Elde Edilen Sonuç

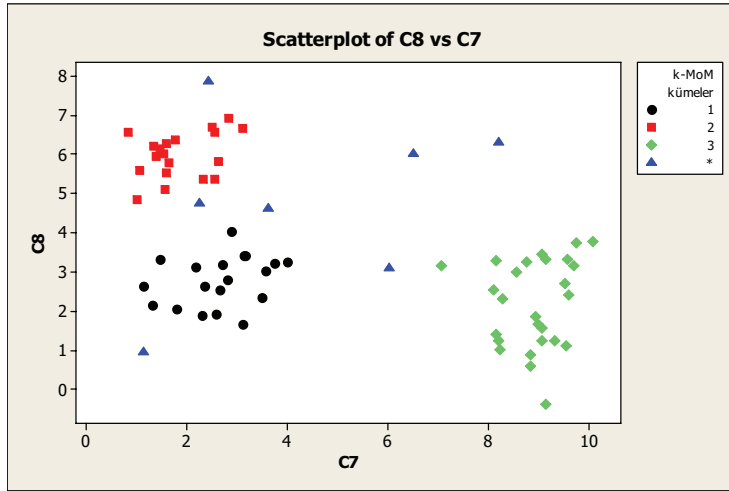


Şekil 15. Sentetik I Veri Setinde  $k=5$  için k-MoM Algoritması ile Elde Edilen Sonuç

Sentetik I veri setinde içerdikleri gözlem sayısı az olan, aykırı değer içeren ve birbirinden aykırık 5 küme bulunmaktadır. Şekil 14 ve Şekil 15 incelendiğinde her iki algoritmanın da genel olarak bu kümeleri tespit etmede başarılı olduğu görülmüştür. Ancak, k-MoM algoritması kümeleri doğru tespit etmesinin yanında veri setinde yer alan aykırı değerleri de belirlemiştir ve  $KİKT=60,3685$  ile k-ortalama yöntemine ait  $KİKT=93,5886$  değerinden daha düşük bir küme içi değişime sahiptir.



Şekil 16. Sentetik II Veri Setinde  $k=3$  için k-ortalama Algoritması ile Elde Edilen Sonuç



Şekil 17. Sentetik II Veri Setinde k=3 için k-MoM Algoritması ile Elde Edilen Sonuç

Sentetik II veri seti, farklı sayıda gözlem içeren ve aykırı değerlerin yer aldığı kümelerden oluşmaktadır. k-ortalama algoritması için Şekil 16, k-MoM algoritması için Şekil 17 incelendiğinde bu veri setinde de algoritmaların aykırı değerleri tespit etmedeki farklılığı gözlenmektedir. k-ortalama algoritmasının KİKT değeri aykırı değerleri tespit edemediği için 143,0664, k-MoM algoritmasının KİKT değeri ise 76,8795 olarak hesaplanmıştır.

#### 4. TARTIŞMA VE SONUÇ

k-ortalama algoritması, veri seti içindeki aykırı değerlerden çok etkilenen bir algoritmadır. Algoritma gereği her nesne mutlaka bir küme içinde yer almaktadır. Bunun sonucu olarak, oluşan kümelerin homojen yapısı bozulmakta ve belirlenen küme merkezleri, küme içinde yer alan aykırı değerlere doğru kaymaktadır. k-MoM algoritması içinde yer alan aykırı değer belirleme adımı nedeniyle veri seti içinde yer alan bir aykırı değer ortaya çıkan bir kümedeki homojenlik yapısını bozması ve merkezi kendine doğru çekmesi engellenmiştir.

Geliştirilen yazılım ile yapılan denemelerde k-MoM algoritmasının küme merkezlerini bulmada ve kümeleri oluşturmada, k-ortalama algoritmasına göre başlangıç noktalarının değişmesinden daha az etkilendiği gözlenmiştir. KİKT değerleri aynı başlangıç noktaları ile uygulama yapıldığında k-ortalama algoritmasına göre her zaman daha küçük çıkmıştır. k-MoM algoritması, tercih edilen küme sayısına bağlı olarak aykırı değer içermeyen veri setlerindeki homojen yapıları, k-ortalama algoritması ile benzer biçimde tespit edebilmesinin yanında, aykırı değer içeren veri setlerine uygulandığında homojen yapıların tespit edilmesi ile beraber aykırı değerleri de belirlemektedir. k-MoM algoritması ve geliştirilen yazılım, aykırı değer içeren veri setlerinde kümeleme çözümlemesi yapılırken, aykırı değerlerden etkilendiği bilinen geleneksel k-ortalama algoritmasına bir seçenek olarak önerilmektedir.

## 5. KAYNAKLAR

- Dallas, E.J., 1998. Applied multivariate methods for data analysts. Duxbury Press.
- Fasulo, D., 1999. An analysis of recent work on clustering algorithms. Technical Report, 01-03-02. Department of Computer Science and Engineering, University of Washington. (in English).
- Guha, S., Rastogi R., Shim, K., 1998. CURE an efficient clustering algorithm for large databases. Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data, L. M. Haas, A. Tiwary (eds.), Seattle, Washington. 73–84.
- Han, J., Kamber M., 2001. Data mining concepts and techniques. Morgan Kauffmann Publishers Inc. San Fransisco.
- Hautamäki, V., Cherednichenko, S., Kärkkäinen, I. Kinnunen, T., Fränti, P., 2005. Improving K-Means by outlier removal. Lecture Notes in Computer Sciences, Springer/Heidelberg. 978–987.
- Huber, P.J., 1964. Robust estimation of location parameters. Annals of Mathematical Statistics, 35, 73–101.
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. University of California Pres, Berkeley.
- Richard, A.J, Dean, W.W., 1992. Applied multivariate statistical analysis. Prentice-Hall, New Jersey.
- Tatlıdil, H., 2002. Uygulamalı çok değişkenli istatistiksel analiz. Ziraat Matbaacılık, Ankara.
- Wilcox R.R., 2003a. Applying contemporary statistical techniques. Academic Pres.
- Wilcox, R.R. 2003b. Multiple comparisons based on a modified one-step M-estimator. Journal of Applied Statistics, 37, 1231–1241.
- Wilcox, R.R., Keselman, H.J., 2003. Repeated measure one-way ANOVA based on a modified one-step M estimator. British Journal of Mathematical and Statistical Psychology, 56, 15–25.
- Zaiane, O., Pei, Y., 2008. <http://www.cs.ualberta.ca/~yaling/Cluster/Applet/Code/Cluster.html>, Temmuz 2008.

## CLUSTER ANALYSIS WITH MODIFIED ONE-STEP M-ESTIMATOR

### ABSTRACT

*Cluster analysis is one of the most widespread multivariate statistical analysis methods. Modified one-step M-estimator is developed with a modification on Huber's M-estimator (Huber, 1964) by Wilcox (Wilcox, 2003a) which is used in analysis of variance (Wilcox and Keselman, 2003) and multiple comparisons (Wilcox, 2003b) since then. In this study, k-means method which is classified as non-hierarchical clustering method has been presented with an algorithm that uses modified one-step M-estimator instead of mean.*

**Keywords:** Modified one-step M-estimator, k-Means clustering, Cluster analysis.