

## **K-MEANS, K-MEDOIDS VE BULANIK C-MEANS ALGORİTMALARININ UYGULAMALI OLARAK PERFORMANSLARININ TESPİTİ**

**Meltem IŞIK\*** , **A. Yılmaz ÇAMURCU\*\***

### **ÖZET**

Kümeleme algoritmalarından bölünmeli kümeleme tekniği, nesnelere giriş parametre sayısı kadar kümeye bölmektedir. Bölünmeli kümeleme algoritmaları, merkez tabanlı kümeleri tespit etmede başarılıdır. Bu çalışmada, başlıca bölünmeli kümeleme algoritmalarından k-means, k-medoids ve bulanık c-means algoritmalarının kümeleme yetenekleri ve performansları karşılaştırılmıştır. Literatürde yer alan sentetik veri setleri kullanılmıştır.

*Anahtar Kelimeler: Bulanık C-means, Bölünmeli Kümeleme, K-means, K-medoids, Kümeleme*

### *APPLIED PERFORMANCE DETERMINATION OF K-MEANS, K-MEDOIDS AND FUZZY C-MEANS ALGORITHMS*

### **ABSTRACT**

Partition based clustering algorithms divide objects to the clusters according to the given input parameter. Partition based clustering algorithms are successful to find center based clusters. In this study, partition based clustering algorithms such as k-means, k-medoids and fuzzy c-means algorithms are compared according to their clustering abilities and performances. Synthetic data sets existing in literature are used in the experiments.

*Keywords: Clustering, Fuzzy C-means, K-means, K-medoids, Partition Based Clustering*

---

\* *Şişli Endüstri Meslek Lisesi, Bilgisayar Bölümü, Şişli, İstanbul*

\*\* *Marmara Üniversitesi, Teknik Eğitim Fakültesi, Elektronik ve Bilgisayar Eğitimi Bölümü, 34722 Göztepe, İstanbul*

## 1. GİRİŞ

Bölünmeli kümeleme algoritmaları,  $k$  giriş parametresini alarak  $n$  tane nesneyi  $k$  tane kümeye böler. Bu teknikler, dendogram gibi iç içe bir kümeleme yapısı üzerinde çalışmak yerine tek-seviyeli kümeleri bulan işlemler gerçekleştirir (Jain vd., 1999). Bütün teknikler merkez noktanın kümeyi temsil etmesi esasına dayanır. Bölünmeli yöntemler, hem uygulanabilirliğinin kolay hem de verimli olması nedeniyle iyi sonuçlar üretirler.

Bu çalışmada, bölünmeli kümeleme algoritmalarından k-means, k-medoids ve bulanık c-means'in performansları uygulamalı olarak karşılaştırılmıştır. Uygulamada sentetik veri setleri kullanılmıştır.

## 2. K-MEANS, K-MEDOİDS VE BULANIK C-MEANS ALGORİTMALARININ YAPILARI

### 2.1. K-means Algoritmasının Yapısı

En eski kümeleme algoritmalarından olan k-means, 1967 yılında J.B. MacQueen tarafından geliştirilmiştir (MacQueen, 1967). En yaygın kullanılan gözetimsiz öğrenme yöntemlerinden birisi olan K-means'in atama mekanizması, her verinin sadece bir kümeye ait olabilmesine izin verir. Bu nedenle, keskin bir kümeleme algoritmasıdır. Merkez noktanın kümeyi temsil etmesi ana fikrine dayalı bir metottur (Han ve Kamber, 2001). Eşit büyüklükte küresel kümeleri bulmaya eğilimlidir.

K-means kümeleme yönteminin değerlendirilmesinde en yaygın olarak karesel hata kriteri SSE kullanılır. En düşük SSE değerine sahip kümeleme sonucu en iyi sonucu verir. Nesnelerin buldukları kümenin merkez noktalarına olan uzaklıklarının karelerinin toplamı (1) nolu eşitlik ile hesaplanmaktadır (Pang-Ning vd., 2006).

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} dist^2(m_i, x) \quad (1)$$

Bu kriterleme sonucu,  $k$  tane kümenin olabildiğince yoğun ve birbirinden ayrı sonuçlanması hedeflenmeye çalışılır. Algoritma, karesel-hata fonksiyonunu azaltacak  $k$  parçayı belirlemeye gayret eder. K-means algoritması, algoritmaya kullanıcı tarafından verilen  $k$  parametresi ile  $n$  tane veriden oluşan veri setini  $k$  adet kümeye böler. Küme benzerliği kümedeki nesnelerin ortalama değeri ile ölçülür, bu da kümenin ağırlık merkezidir (Xu ve Wunsch, 2005).

## 2.2. K-medoids Algoritmasının Yapısı

K-medoids algoritmasının temeli, verinin çeşitli yapısal özelliklerini temsil eden  $k$  tane temsilci nesneyi bulma esasına dayanır (Kaufman ve Rousseeuw, 1987). Temsilci nesne medoid olarak adlandırılır ve kümenin merkezine en yakın noktadır. Bir grup nesneyi  $k$  tane kümeye bölerken asıl amaç, birbirine çok benzeyen nesnelerin bir arada bulunduğu ve farklı kümelerdeki nesnelerin birbirinden benzersiz olduğu kümeleri bulmaktır. En yaygın kullanılan k-medoids algoritması, 1987 yılında Kaufman and Rousseeuw tarafından geliştirilmiştir (Kaufman ve Rousseeuw, 1990). Temsilci nesne, diğer nesnelere olan ortalama uzaklığı minimum yapan kümenin en merkezi nesnesidir. Bu nedenle, bu bölünme metodu her bir nesne ve onun referans noktası arasındaki benzersizliklerin toplamını küçültme mantığı esas alınarak uygulanır. Kümeleme literatüründe temsilci nesnelere çoğunlukla merkez tipler (centrotypes) denilmektedir. PAM (Partitioning Around Medoids) algoritmasında temsilci nesnelere medoid olarak adlandırılmaktadır (Kaufman ve Rousseeuw, 1990). Amacın  $k$  tane nesneyi bulmak olmasından dolayı, k-medoids metodu olarak adlandırılmaktadır.  $k$  adet temsilci nesne tespit edildikten sonra her bir nesne en yakın olduğu temsilciye atanarak  $k$  tane küme oluşturulur. Sonraki adımlarda her bir temsilci nesne temsilci olmayan nesne ile değiştirilerek kümelemenin kalitesi yükseltinceye kadar ötelenir. Bu kalite nesne ile ait olduğu kümenin temsilci nesnesi arasındaki ortalama benzersizlik maliyet fonksiyonu kullanılarak değerlendirilir.

## 2.3. Bulanık c-means Algoritmasının Yapısı

Bulanık c-means (FCM) algoritması, bulanık bölünmeli kümeleme tekniklerinden en iyi bilinen ve yaygın kullanılan yöntemdir. Bulanık c-means algoritması 1973 yılında Dunn tarafından ortaya atılmış ve 1981’ de Bezdek tarafından geliştirilmiştir (Höppner vd., 2000). Bulanık c-means algoritması da amaç fonksiyonu temelli bir metottur. Bulanık c-means metodu, nesnelerin 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. Amaç fonksiyonun belirlenen minimum ilerleme değerine yakınsaklaşmasıyla kümeleme işlemi tamamlanır.

Algoritma, en küçük kareler yönteminin genellemesi olan aşağıdaki amaç fonksiyonunu öteleyerek minimize etmek için çalışır (Höppner vd., 2000).

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty \quad (2)$$

U üyelik matrisi rasgele atanarak algoritma başlatılır. İkinci adımda ise merkez vektörleri hesaplanır. Merkezler (3) nolu eşitlik ile hesaplanır (Höppner vd., 2000).

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m} \quad (3)$$

Hesaplanan küme merkezlerine göre, U matrisi (4) nolu eşitlik kullanılarak yeniden hesaplanır. Eski U matrisi ile yeni U matrisi karşılaştırılır ve fark  $\epsilon$ 'dan küçük olana kadar işlemler devam eder (Moertini, 2002).

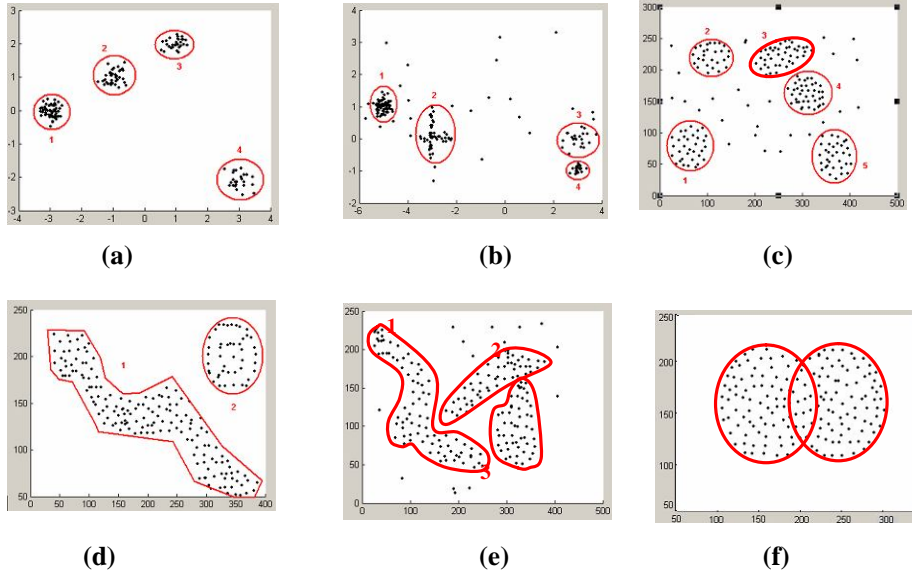
$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_i - c_i\|}{\|x_i - c_k\|} \right)^{2/(m-1)}} \quad (4)$$

Kümeleme işlemi sonucunda bulanık değerler içeren U üyelik matrisi kümelemenin sonucunu yansıtır. İstenirse, berraklaştırma yapılarak bu değerler yuvarlanıp 0 ve 1'lere dönüştürülebilir.

### 3. KULLANILAN YAZILIM ORTAMI VE VERİ SETLERİ BÖLÜNME Lİ KÜMELEME UYGULAMA PROGRAMININ ARAYÜZ TASARIMI

Bu çalışmada, yazılımların geliştirilmesinde literatürde yapılmış çalışmalardan ve MATLAB'te bulunan hazır fonksiyonlardan da yararlanılmıştır. K-means algoritmasının uygulanmasında Kadri Teknomo tarafından oluşturulan k-means fonksiyonu (Teknomo, 2005) örnek alınmıştır. Burada tarafımızca bazı kodlarda düzenlemeler yapılmıştır. K-medoids algoritması geliştirilmeden önce (Salem ve Nandi, 2005)'da verilen matematiksel fonksiyonlar ile (Zaiane ve Pei, 2005)'da verilen sözde kodlar incelenerek K-medoids algoritmasının kodları oluşturulmuştur. Bulanık c-means algoritmasında MATLAB'ın bulanık toolbox'ında bulunan Roger Jang tarafından geliştirilen hazır fonksiyonlar kullanılmıştır.

Uygulamalarımızda kullanılan veri setleri Şekil 1'de görülmektedir. Örüntüler ve küçük biçimler (Patterns ve Small Shapes) sentetik veri setleri, Finlandiya Uppsala Üniversitesi Bilgisayar Bilimleri Bölümünden Profesör Tobias Lindhal ve Per Gustafsson tarafından hazırlanan veritabanından alınmıştır (Lindahl ve Gustafsson, 2005). Documents\_Sim, Mars ve Image Extraction sentetik veri setleri, Kanada Alberta Üniversitesi Bilgisayar Bilimleri Bölümünde Osmar Zaiane ve Yaling Pei tarafından hazırlanmış veritabanlarından alınmıştır (Zaiane ve Pei, 2005).



**Şekil 1. Sentetik Veri Setleri: (a) Örüntüler, (b) Küçük Biçimler, (c) Documents\_Sim, (d) Mars, (e) Image Extraction, (f) Yuvarlaklar**

Şekil 1.a'daki örüntüler veri seti 150 adet noktanın x ve y koordinatlarını içermektedir. Bu veri setinde yoğun, sıra dışı veriler içermeyen, birbirinden iyi ayrılmış dört adet küme bulunmaktadır. Şekil 1.b'deki küçük şekiller [200\*2] veri seti ise dört çeşit küme dağılımını temsil eden bir yapıya sahiptir. Bu kümeler sırasıyla aşağıdaki gibidir:

1. Yoğun ve sıradışı nesnelere içeren
2. Küresel olmayan ve sıradışı nesnelere içeren
3. Seyrek
4. Yoğun ve sıradışı nesnelere içermeyen

Şekil 1.c' deki Documents\_Sim [200\*2] veri seti sıra dışı veriler içeren beş adet küresel kümeden oluşmaktadır. Mars [200\*2] veri seti farklı şekillerde kümelere ayrılan ayrıca sıra dışı veriler içermektedir. Image Extraction [200\*2] veri setinde şekilleri ve büyüklükleri birbirinden oldukça farklı iki küme bulunmaktadır. Tarafımızca hazırlanan Yuvarlaklar [176\*2] sentetik veri seti ise, birbirine örtüşen iki küresel küme içeren bir yapıya sahiptir.

#### 4. SONUÇLAR VE KARŞILAŞTIRMA

k-means, k-medoids ve bulanık c-means algoritmaları, Şekil 1’de görülen sentetik veri setleri ile gerçekleştirilen uygulamaları ve elde edilen kümeleme sonuçları burada açıklanmaktadır.

##### 4.1. Algoritmaların Küresel Kümeler İçeren Veri Setlerinde Uygulanması

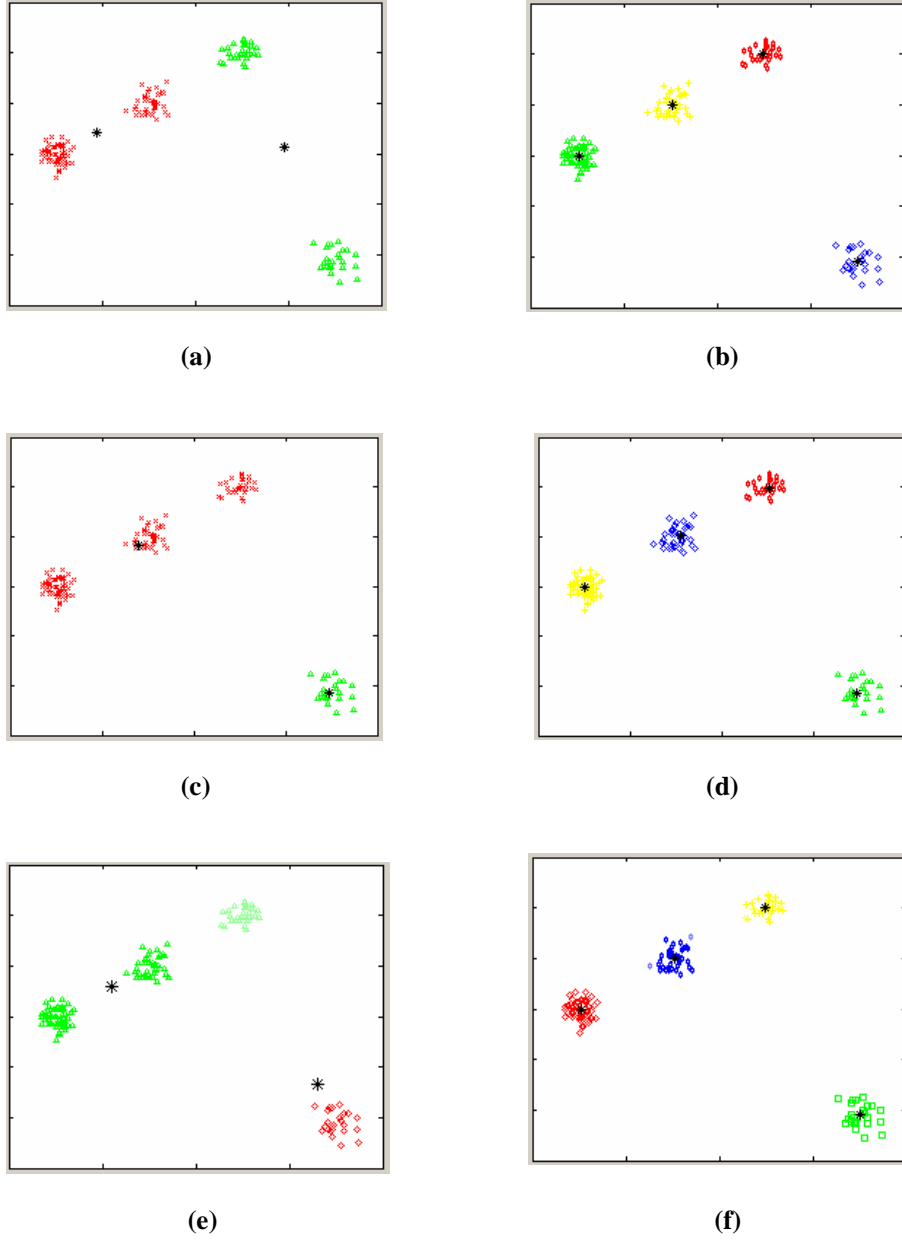
K-means algoritmasının örüntüler veri seti üzerinde farklı k değerleri için gerçekleştirilen uygulama sonuçları Şekil 2’de görülmektedir. Küme merkezleri “\*” işareti ile gösterilmektedir. Şekil 2.a’da görüldüğü gibi k=2 değeri için 1. ve 2. küreleri oluşturan noktaların tamamı 1. kümeye, 3. ve 4. küreleri oluşturan noktaların büyük tamamı 2. kümeye girmiştir.

Şekil 2.d’de görüldüğü gibi k-medoids algoritması k=4 değeri için k-means ile aynı sonucu üretmiştir. Fakat k=2 için iki algoritmanın sonuçları farklıdır. K-medoids 1., 2., ve 3. küreleri aynı kümeye 4. küreyi ise ayrı bir kümeye dahil etmiştir. Şekil 2 e ve f’de bulanık c-means algoritmasının sonuçları görülmektedir. Bu sonuçlara göre Bulanık c-means algoritması da k=2 için ilk 3 küreyi gruplamış 4. küreyi ise diğer kümeye koymuştur.

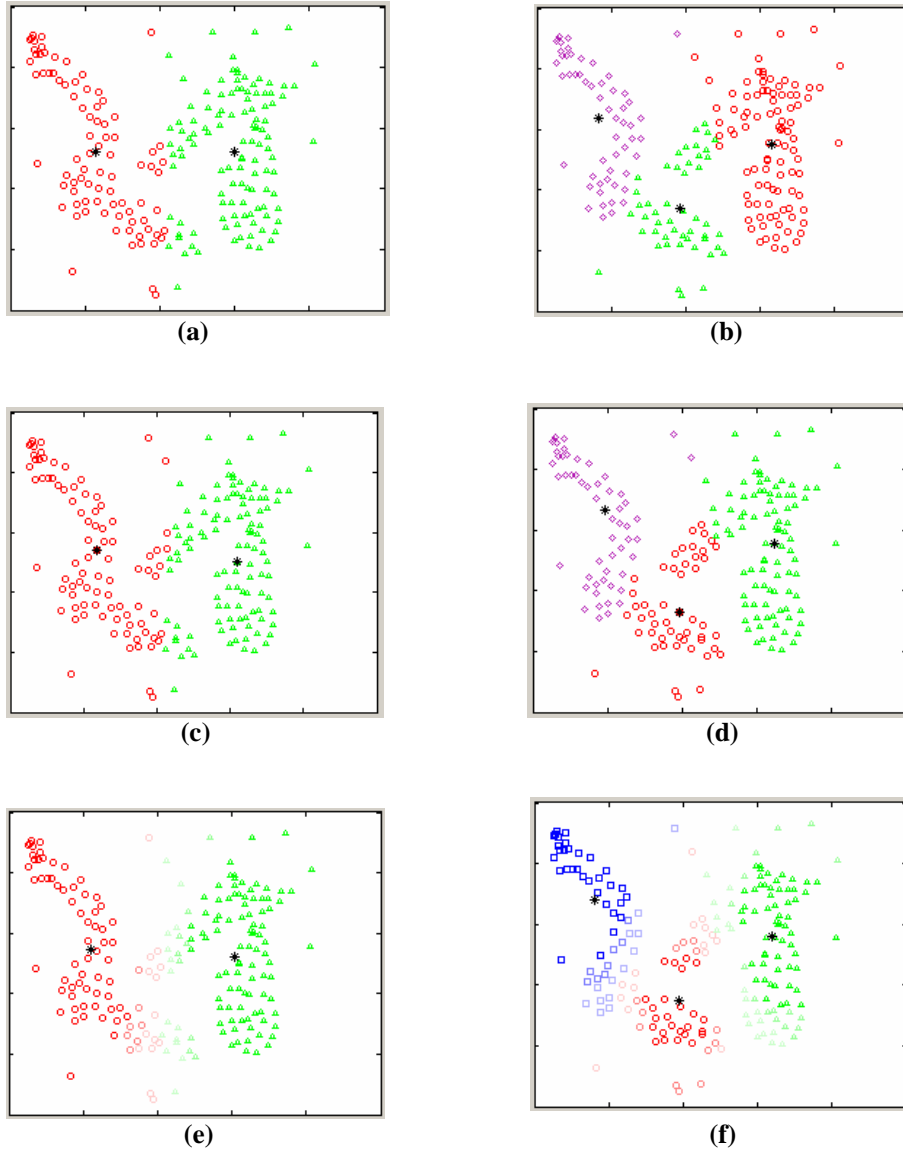
Bütün algoritmalar k=4 için ideal bir kümeleme gerçekleştirmiştir. Birbirinden iyi ayrılmış, küresel ve yoğun kümeler içeren veri setlerinde algoritmaların üçü de iyi sonuç vermektedir. Kümelerin dağılımı ve elemanları arasında bir fark görülmemektedir. Bu kümeleme işleminde fark k-means’in en çabuk, k-medoids’in ise en geç cevap vermesidir. Fakat, k=2 için aynı sonuçlar elde edilmemiştir. Veri setinde ilk üç küre birbirine yakın olup dördüncü küre ise daha uzakta bulunmaktadır. Bulanık c-means ve k-medoids yoğun bölgeyi tespit etmeyi başarırken k-means farklı sonuç vermiştir. Bu nedenle, bulanık c-means ve k-medoids’in, k-means’e göre kümeleri daha iyi tespit ettiği anlaşılmaktadır. Ayrıca k-medoids temsilci nesne kullanması nedeniyle küme merkezlerini daha iyi tespit etmiştir. Bu örnekte k-medoids’in kümeleme sonucunun en başarılı olduğu görülür.

##### 4.2. Bölünmeli Kümeleme Algoritmalarının Küresel Olmayan Kümeler İçeren Veri Setlerinde Uygulanması

Mars veri seti ise hem küresel olmayan hem de sınırları tam belirgin olmayan kümeler içeren bir veri setidir. k=2 ve k=3 parametreleri için uygulamalarda Şekil 3’de görüldüğü gibi tüm sonuçlar başarısızdır.



Şekil 2. Örüntüler Veri Setinde,  $k=2$  ve  $k=3$  Parametreleri İçin (a, b) k-means, (c, d) k-medoids İçin, (e, f) Bulanık c-means Algoritması ile Elde Edilen Sonuçlar

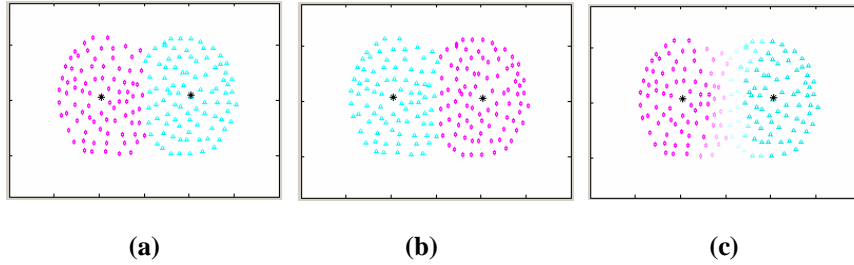


Şekil 3. Mars Veri Setinde,  $k=2$  ve  $k=3$  Parametreleri İçin (a, b) k-means, (c, d) k-medoids İçin, (e, f) Bulanık c-means Algoritması İle Elde Edilen Sonuçlar



#### 4.3. Bölünmeli Kümeleme Algoritmalarının Birbiriyle Örtüşen Kümeler İçeren Veri Setlerinde Uygulanması

Yuvarlaklar veri seti örtüşen iki küme içeren bir veri setidir. Şekil 4’de k-means, k-medoids ve bulanık c-means algoritmalarının sonuçları görülmektedir. Bu veri seti üzerinde  $k=2$  parametresi için uygulama yapılmıştır. K-means ve k-medoids algoritmaları keskin yöntemler olduğu için bir nesnenin sadece bir kümeye ait olmasına izin vermiştir. Ancak örtüşen kümeler içeren veri setlerinde bazı nesnelere birden fazla kümeye ait olmaktadır ve kümelerin kesişim bölgeleri keskin sınırlarla ayıramaz. Bu nedenle, nesnelere her kümeye üyelik değerinin incelenmesi gerekir. Bulanık c-means algoritması küme içindeki üyelik değeri yüksek olan nesnelere koyu renkle, diğer kümeye üyeliği bulunduğu için üyelik değeri düşük olan nesnelere ise açık renkle belirtmiştir. Böylece iki kümenin kesişim bölgesindeki her iki kümeye üyeliği olan nesnelere doğru ifade edilmiştir. Bulanık c-means örtüşen kümeleri bulmada başarılı olmuş ancak k-means ve k-medoids algoritmaları başarısız olmuştur.



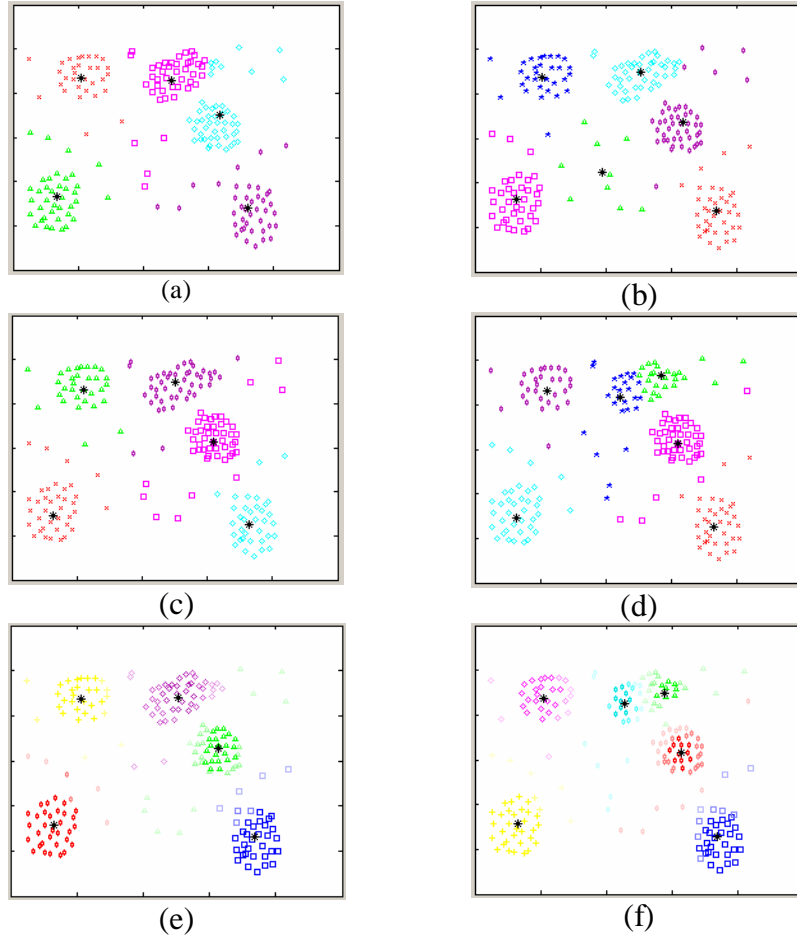
Şekil 4. Yuvarlaklar Veri Setinde, (a) k-means (b) k-medoids Algoritması İle Elde Edilen Sonuçlar (c) Bulanık c-means Algoritması İle Elde Edilen Sonuçlar

Document\_sim veri seti üzerinde  $k=5$  ve  $k=6$  parametreleri için yapılan K-means, K-medoids ve bulanık c-means algoritmalarının kümeleme analizi sonuçları Şekil 5’de görülmektedir.

Document\_sim veri seti, örüntüler veri seti gibi küresel kümelerden oluşmasına karşın sıra dışı veriler içerdiği için algoritmaların hepsi de başarılı olmamıştır. K-medoids ve bulanık c-means algoritmaları sıra dışı verilerden etkilenmemiş ve  $k=5$  değeri için yapılan her testte yoğun bölgeleri doğru olarak tespit etmiştir ancak k-means algoritması sıra dışı verilerden oldukça etkilenmiştir. K-means algoritmasında küme merkezleri nesnelere aritmetik ortalaması hesaplanarak tespit edildiği için sıra dışı verilerin küme merkezi değerinde oldukça ağırlıklı etkisi olur. K-means uygulamasında küme merkezlerinin sıra dışı verilere doğru ilerlemiş olduğu görülmektedir. Ayrıca rasgele belirlenen başlangıç değerleri sıra dışı

verilerin etkisini deđiřtirdiđi için k-means'le yapılan her testte ayrı sonuçlar elde edilmiřtir.

k=6 için kümeleme sonucu incelendiđinde k-means'in sıra dıřı verilerden olumsuz etkileniři daha net görölmektedir. K-medoids ve bulanık c-means yine daha yođun kümeleri tespit edebilmiř, ancak k-means altıncı kümeyi sıra dıřı verilerden oluřturmuřtur.



řekil 5. Document\_Sim Veri Setinde, k=5 ve k=6 Parametreleri İin, (a, b) k-means, (c, d)k-medoids İin, (e, f) Bulanık c-means Algoritması İle Elde Edilen Sonular

#### 4.4. Bölünmeli Kümeleme Algoritmalarının Büyüklükleri Farklı Kümeler Üzerinde Uygulanması ve Küme Sayısının Etkisi

Image extraction veri seti üzerinde  $k=2$  ve  $k=3$  parametreleri için yapılan K-means, K-medoids ve bulanık c-means algoritmalarının kümeleme analizi sonuçları Şekil 6'da görülmektedir. Image extraction veri setinde  $k=2$  için bir tane küresel bir tane de ince uzun bir küme bulunması gerekir. Küresel küme diğer kümenin yaklaşık dörtte biri kadardır. Üç algoritma da nesnelere arasındaki bağlantıyı inceleyerek kümeleme yapamadığı için büyüklükleri farklı kümeleri tespit edememekte, merkez noktalara göre nesnelere kümelediği için eşit büyüklükteki kümeleri bulma eğilimi göstermektedir. Bu nedenle, şekillerden anlaşılacağı gibi üç metot da gerçek kümeleri bulmayı başaramamıştır.

Görüldüğü gibi k-means, k-medoids ve bulanık c-means algoritmaları k sayısına bağımlı sonuçlar üretmektedir. Çünkü Image extraction veri setinde iki adet küme bulunmasına karşın,  $k=3$  parametresi nedeniyle büyük küme ikiye ayrılarak veri seti üç kümeye bölünmektedir. Buradan da, algoritmaların doğru k değerini tespit etme yeteneğinin olmadığı görülmektedir.

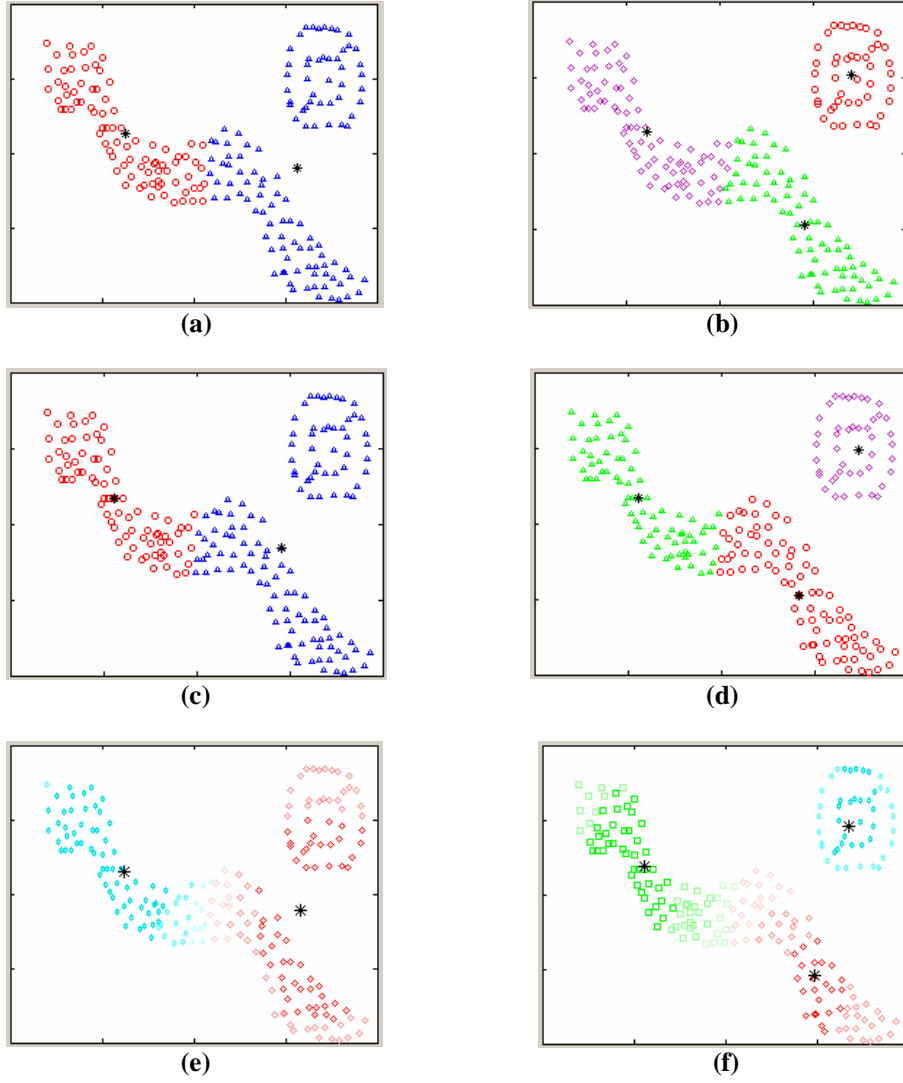
#### 4.5. Başlangıç Değerlerinin Algoritmalar Üzerindeki Etkisi

Başlangıç değerlerinin kümeleme başarısındaki etkisi için yapılan testlere ilişkin K-means, K-medoids ve bulanık c-means algoritmasının sonuçları Şekil 7'de görülmektedir. Küçük şekiller veri seti üzerinde  $k=4$  için çok sayıda test yapılmıştır. Her algoritma için en çok elde edilen ikişer sonuç şeklinde görülmektedir. K-medoids algoritması, bütün testlerde doğru kümeleri tespit etmiştir. Bu nedenle, başlangıç değerlerinden etkilenmediği, kararlı bir yapıya sahip olduğu görülmüştür. Bulanık c-means algoritması, yapılan testlerde iki farklı sonuç üretmiştir. Bulanık c-means'nin genelde kararlı bir yapıya sahip olduğu, ancak başlangıç değerlerinden biraz etkilendiği görülmüştür. K-means algoritması ise birçok testte farklı sonuçlar üretmiş ve başlangıç değerlerine çok duyarlı olduğu görülmüştür.

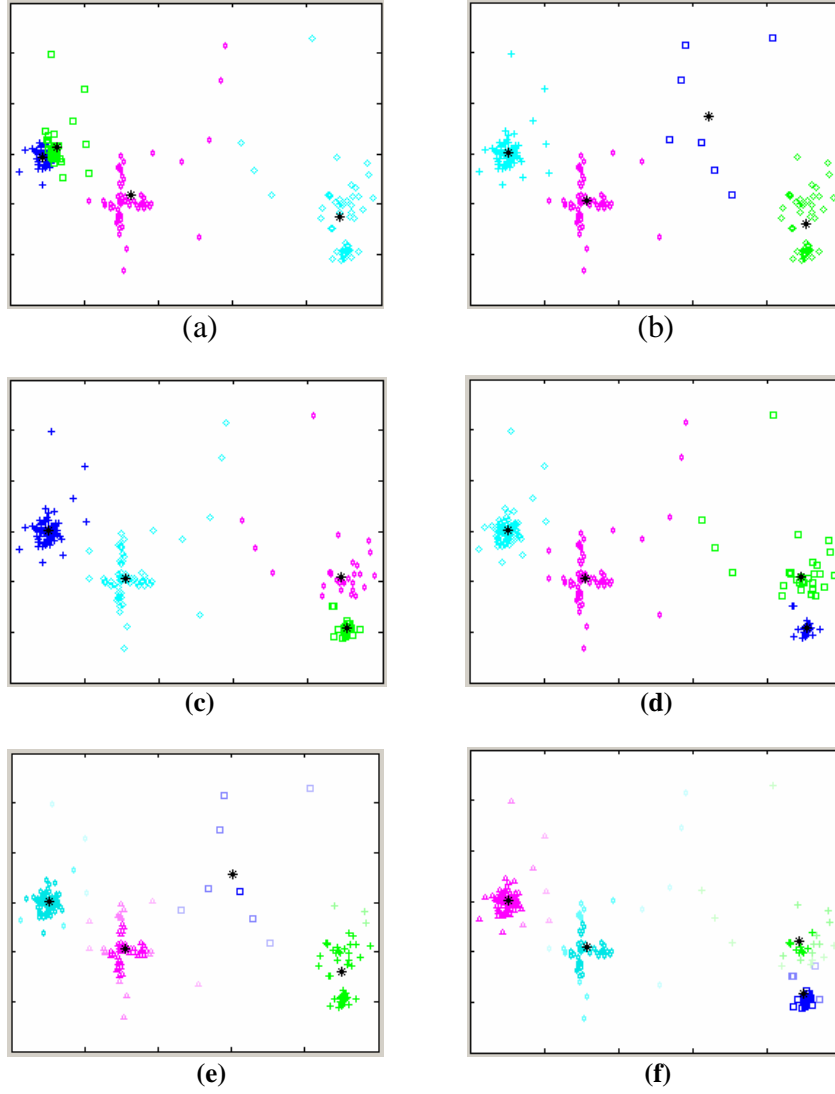
### 5. TARTIŞMA

K-means, k-medoids ve bulanık c-means algoritmalarının sentetik veri setleri üzerinde uygulanması ile elde edilen sonuçların karşılaştırması aşağıda verilmektedir.

Bölünmeli kümeleme algoritmaları, nesnelere küme merkezine olan uzaklığını inceleyerek kümeleme işlemi yaptığı için küresel kümeleri bulmaya eğilimlidir. Yapılan testlerde, algoritmaların küresel kümeleri bulmaya eğilimli olduğu bütün örneklerde görülmüştür.



**Şekil 6. Image Extraction Veri Setinde, k=2 ve k=3 Parametreleri İçin, (a, b) k-means, (c, d) k-medoids İçin, (e, f) Bulanık c-means Algoritması İle Elde Edilen Sonuçlar**



**Şekil 7. Küçük Şekiller Veri Setinde, (a, b) k-means Algoritması, (c, d) k-medoids Algoritması, (e, f) c-means Algoritması İle Elde Edilen Sonuçlar**

Üç algoritmanın da uygun bir değer olmasa bile verilen k sayısına bağlı kaldığı ve farklı şekillerdeki kümeleri tespit etmede yetenekli olmadığı da gözlemlenmiştir. Kümeleri ve küme merkezlerini bulmada en başarılı algoritma k-medoids'dir. Kümelerin üyelik matrisini hesaplaması nedeniyle bulanık c-means algoritması sonucun farklı şekillerde yorumlanmasına izin vermektedir. Örtüşen kümeleri tespit

etmede bulanık c-means'in başarılı, k-means ve k-medoids'in ise başarısız olduğu görülmüştür. Bulanık c-means ve k-means algoritmaları, aynı kümeye aynı k sayısı ile yapılan testlerde farklı sonuçlar üretebilmektedir.

Bulanık c-means'in, k-means'e göre başlangıç değerlerinden daha az etkilendiği genellikle daha kararlı sonuçlar ürettiği gözlemlenmiştir. K-medoids algoritmasının ise başlangıç değerinden etkilenmediği sonuçlarının daha kararlı olduğu yapılan testlerde görülmüştür. K-medoids algoritmasının sıra dışı verilerden hiç etkilenmediği, bulanık c-means'nin ise küçük bir oranda etkilendiği, k-means'in ise çok etkilendiği görülmüştür.

Zaman açısından en avantajlı algoritma ise k-means algoritması olup, bulanık c-means k-means'e göre biraz yavaş, k-medoids algoritması ise çok yavaş çalışmaktadır. Bu nedenle klasik k-medoids algoritması küçük veri setleri için uygun bir algoritmadır.

## 6. DEĞERLENDİRME

Bu çalışmada, bölünmeli kümeleme algoritmalarından k-means, bulanık c-means ve k-medoids algoritmalarının sentetik veri setlerindeki performansları araştırıldı. Üç algoritmanın da seçilen veri kümesini istenilen sayıda kümeye böldükleri gözlemlendi. Bu üç algortmadan en iyi kümeleme işlemini gerçekleyen k-medoids algoritmasıdır, ancak k-medoids çok boyutlu verilerde ve çok büyük veri kümelerinde zaman ve bellek yetersizliği gibi ciddi problemlere neden olmaktadır.

Üç algortmada da başlangıç küme merkezleri rastgele atanmaktadır. Bu rasgele atama sonucuna k-means algoritmasında sonuçlar, k-medoids ve bulanık c-means algoritmasına göre daha çok değişkenlikler göstermektedir.

Üç algortmada da nesnelere kümelemek için Öklid uzaklık ölçütü kullanılmış olduğundan, algoritmaların küresel kümeleri bulma eğilimli olduğu gözlemlenmiştir. Bulanık c-means algoritmasının diğer iki algortmadan ayrılan en belirgin özelliği, nesnelere aitlik ifadesidir. Diğer iki algortmada nesnelere sadece bir kümeyle ait olup, diğer kümelerle aitlik oranı sıfırdır. Bulanık c-means algoritmasında ise bir nesnenin her kümeyle aitlik oranı belirlenir ve nesnenin bütün kümelerle aitlik oranının toplamı birdir.

## 7. KAYNAKÇA

Han, J., and Kamber, M., (2006), Data Mining Concepts and Techniques, Morgan Kaufmann Publishers Inc.

Höppner, F., Klawonn, F., Kruse, R., and Runkler, T., (2000), Fuzzy Cluster Analysis, John Wiley&Sons, Chichester.

- Jain, A. K., Murty, M. N., and Flynn, P. J., (1999), "Data Clustering: A Review", *ACM Computing Surveys*, 31, 3.
- Kaufman, L., and Rousseeuw, P. J., (1987), Clustering by Means of Medoids," *Statistical Data Analysis Based on The L1-Norm and Related Methods*, edited by Y. Dodge, North-Holland, 405–416.
- Kaufman, L., and Rousseeuw, P. J., (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley and Sons.
- Lindhahl, T., and Gustaffson, P., "Sentetik Veri Seti, Küçük Şekiller ve Örüntülerin Kaynağı", <http://user.it.uu.se/~kostis/Teaching/DM/Assignments/> (Erişim Tarihi: Mayıs 2005)
- MacQueen, J. B., (1967), MacQueen, Some Methods for Classification and Analysis of Multivariate Observations, *Proc. Symp. Math. Statist. and Probability (5th)*, 281–297.
- Moertini, V. S., (2002), "Introduction to Five Clustering Algorithms", *Integral*, 7, 2.
- Pang-Ning Tan, P. N., Steinbach, M., and Kumar, V., (2006), *Introduction to Data Mining*, Addison Wesley.
- Salem, S. A., and Nandi, A. K., (2005), "New Assessment Criteria for Clustering Algorithms", *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP-2005)*, Mystic, CT, USA, 285-290.
- Teknomo, K., <http://people.revoledu.com/kardi/tutorial/kMean/>, (Erişim Tarihi: Eylül 2005)
- Xu, R., and Wunsch, II. D., (2005), "Survey of Clustering Algorithms", *IEEE Transactions On Neural Networks*, 16, 3.
- Zaiane, O., and Pei, Y., Sentetik Veri Seti Documents\_Sim, Mars ve Image Extraction'in kaynağı: <http://www.cs.ualberta.ca/~yaling/Cluster/Applet/Code/Cluster.html> (Erişim Tarihi: Eylül 2005).