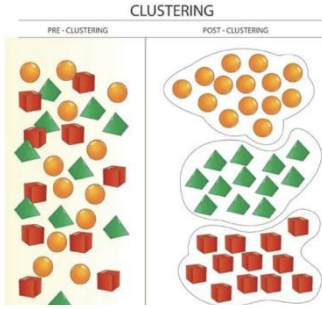


## ÖDEV 3

Selen Erdoğan  
Selenerdogan2019@gtu.edu.tr  
Elektronik Mühendisliği Bölümü, GTÜ

### I. GİRİŞ

K-Means algoritması, veri madenciliği ve veri analitiği gibi alanlarda kullanılan bir kümeleme (clustering) algoritmasıdır. Bu algoritma, benzer özelliklere sahip veri noktalarını bir araya getirerek verileri gruplara ayırmak için kullanılır. K-Means algoritması, özellikle veri setinin yapısını anlamak, veri analizi yapmak veya veri madenciliği uygulamalarında kullanılır.



Şekil 1

### II. YÖNTEM

#### A. K-means Algoritması

1-Başlangıçta, kullanıcı tarafından belirtilen K sayısına (küme sayısı) göre rastgele K tane merkez nokta seçilir. Bu merkez noktalar, veri setindeki örneklerden rastgele seçilir.

2-Her veri noktası, en yakın merkeze atanır. Bu atama işlemi, her veri noktasının en yakın merkeze olan uzaklığına göre hesaplanır. Bu sayede her veri noktası bir kümenin üyesi olur.

3-Her kümeye atanmış veri noktalarının ortalaması alınarak yeni merkezler hesaplanır. Bu yeni merkezler, kümelerin kütle merkezlerini temsil eder.

4-Yeni merkezler hesaplandıktan sonra, her veri noktası tekrar en yakın merkeze atanır. Bu işlem, veri noktalarının yeniden gruplandırılmasını sağlar.

5-Yeni gruplandırma sonucunda merkezler tekrar hesaplanır, ve bu adımlar birkaç iterasyon boyunca tekrarlanır. İterasyonlar, merkezler artık önemli bir değişiklik göstermediğinde veya belirli bir iterasyon sayısına ulaşıldığında sona erer.

6-Algoritma sonlandığında, veri noktaları K kümesine ayrılmış olur. Her veri noktası, en yakın merkeze olan uzaklığına göre bir kümenin üyesi olarak kabul edilir.

#### B. Salinas Haritalarına K-means Algoritması Entegre Etme

Haritaları daha iyi analiz edebilmek için k-means algoritması verilen veri setlerine uygulandı. Şekil 2 ve 3' te elde edilen sonuçlar paylaşılmıştır.

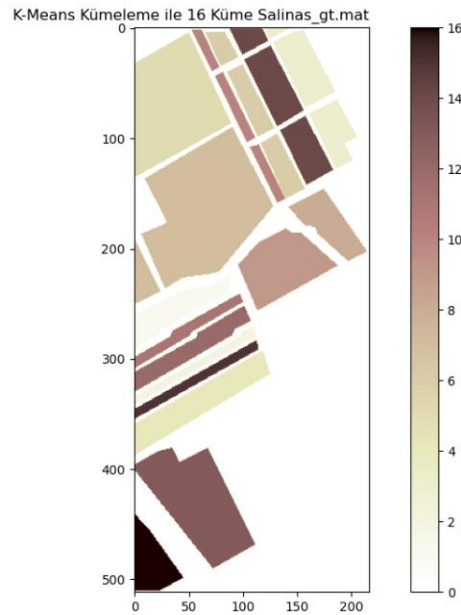
Başlangıç merkezinin seçimi için rastgele bir başlangıç merkezi seçildi. Her veri noktası için mevcut merkezlere olan uzaklık hesaplandı.

Uzaklıklar üzerinden bir olasılık dağılımı oluşturuldu ve bu dağılım kullanılarak sonraki merkez seçildi. Bu süreç, toplam k merkez elde edilene kadar tekrarlandı.

Başlangıç merkezlerinin daha dengeli bir şekilde dağıtılmasını sağlanarak, K-Means algoritmasının performansını iyileştirmek amaçlandı. Başlangıç merkezleri seçildi. Her veri noktası için, tüm merkezlere olan mutlak uzaklıklar hesaplandı. Her veri noktası en yakın merkeze atandı. Her küme için yeni merkezler, kümeye atanmış noktaların ortalaması olarak hesaplandı.

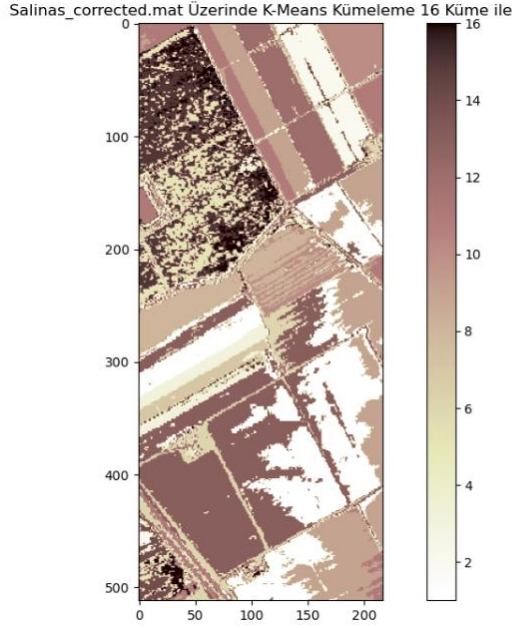
Eğer yeni merkezler eski merkezlerden belirli bir eşik değerden daha az farklılık gösterirse ya da maksimum iterasyon sayısına ulaşırsa, döngü sonlandırıldı.

Merkezlerin yeni ve eski pozisyonları arasındaki toplam uzaklık, bir eşik değeri ile karşılaştırıldı. Bu eşik değeri aşılmazsa, algoritma yakınsamış kabul edildi ve döngü sonlandırıldı.



Şekil 2

# ELM472 Makine Öğrenmesinin Temelleri

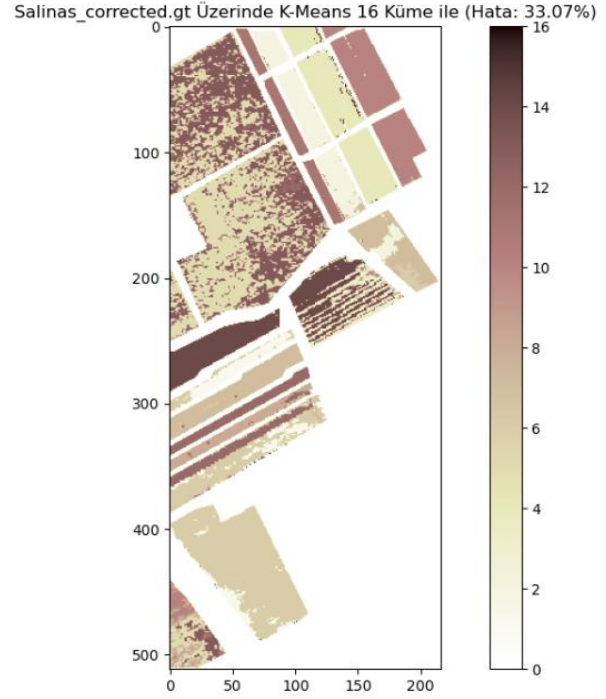


Şekil 3

## III. HATA FONKSİYONU

Hata bulma fonksiyonu ile K-Means kümeleme algoritması sonuçlarının doğruluğunu değerlendirmek amaçlandı. Bu fonksiyon ile, kümeleme sonuçlarını gerçek zemin gerçeği verileriyle karşılaştırarak, algoritmanın performansını ölçüldü. Fonksiyon, önce kume\_sayisi x kume\_sayisi boyutlarında bir karışıklık matrisi oluşturdu. Bu matris, algoritmanın her bir kümesini gerçek değerlerle karşılaştırdı. Elde edilen eşleştirme indeksleri kullanılarak, algoritmanın kümeleri gerçek kümelerle eşlendi. Bu, algoritmanın küme etiketlerini gerçek küme etiketleriyle uyumlu hale getirir. Ayarlanmış küme etiketleri, gerçek küme etiketleriyle karşılaştırıldı ve her iki etiketin eşleşmediği yerler sayıldı. Eşleşmeyen etiketlerin oranı hesaplanarak, algoritmanın toplam hata yüzdesi bulundu.

Bu fonksiyon, kümeleme algoritmasının ne kadar doğru çalıştığını anlamak için yazıldı.



Şekil 4

## IV. SONUÇLAR

Hiperspektral görüntü verileri üzerinde K-Means kümeleme algoritmasını uygulanarak, verileri benzer özelliklere sahip kümelerle ayrıldı. Başlangıç merkezlerini daha dengeli bir şekilde seçerek, K-Means kümelemenin daha etkili olmasını sağlandı. Bu, rastgele başlangıç noktalarının neden olabileceği kötü kümeleme sorunlarını azaltılmasına sebep oldu. Kümeleme sonuçlarının gerçek verilerle uyumunu değerlendirmek için ayarlanmış hata oranı hesaplandı. Kümeleme sonuçları görselleştirilerek, farklı kümelerin mekansal dağılımı gösterildi.

## KAYNAKÇA

<https://scikitlearn.org/stable/modules/generated/sklearn.cluster.KMeans.html#:~:text='k%2Dmeans%2B%2B'%20%3A,This%20technique%20speeds%20up%20convergence.>

[https://tr.wikipedia.org/wiki/K-means\\_k%C3%BCmeleme](https://tr.wikipedia.org/wiki/K-means_k%C3%BCmeleme)