

## YAZ 104 Temel Programlama II

Bahar 2020

### Mini Proje 3

(80 puan)

28 Nisan 2020

Bu mini projede Kuzey Amerika'daki eyaletlerin, yine Kuzey Amerika'da yetişen çeşitli bitki türleri açısından gruplara ayıracağız, yani eyaletler arasında kümeleme gerçekleştireceğiz. Kullanıcı, çeşitli bitkilerin Latince'deki isimleri ve hangi eyaletlerde doğal olarak yetiştiklerini listeleyen bir veri setini yükleyerek programınızı kullanmaya başlamalı. Bitki veri seti diye adlandırabileceğimiz bu text dosyanın formatı şu şekildedir:

*Bitkinin Latince ismi, eyalet\_kisaltması1, state\_ eyalet\_kisaltması2, state\_ eyalet\_kisaltması3, ....*

Veri seti dosyasında aynı bitki, adındaki bazı eklemelerle pek çok kez geçmektedir. Her geçişinde ilk kelime bir bitkinin farklı isimleri için aynı olmakla birlikte diğer kelimelerde farklılıklar bulunmaktadır. Örneğin aşağıdaki satırların tamamı **abies** isimli bitkinin doğal olarak yetiştiği eyaletlerin bilgilerini vermekle birlikte, her satırda aynı bitkinin ismi **abies** kelimesiyle başlayıp farklı sayıdaki değişik kelimelerle devam etmektedir. Bütün bu satırlarda verilen bilgilerin **abies** isimli tek bir bitkinin varyasyonlarına ait olduğu kabul edilecektir. Her satırda ilk virgül karakterinden sonra o bitkinin yetiştiği eyalet kısaltmaları, yine virgül karakteriyle ayrılmış bir şekilde, sıralanmaktadır.

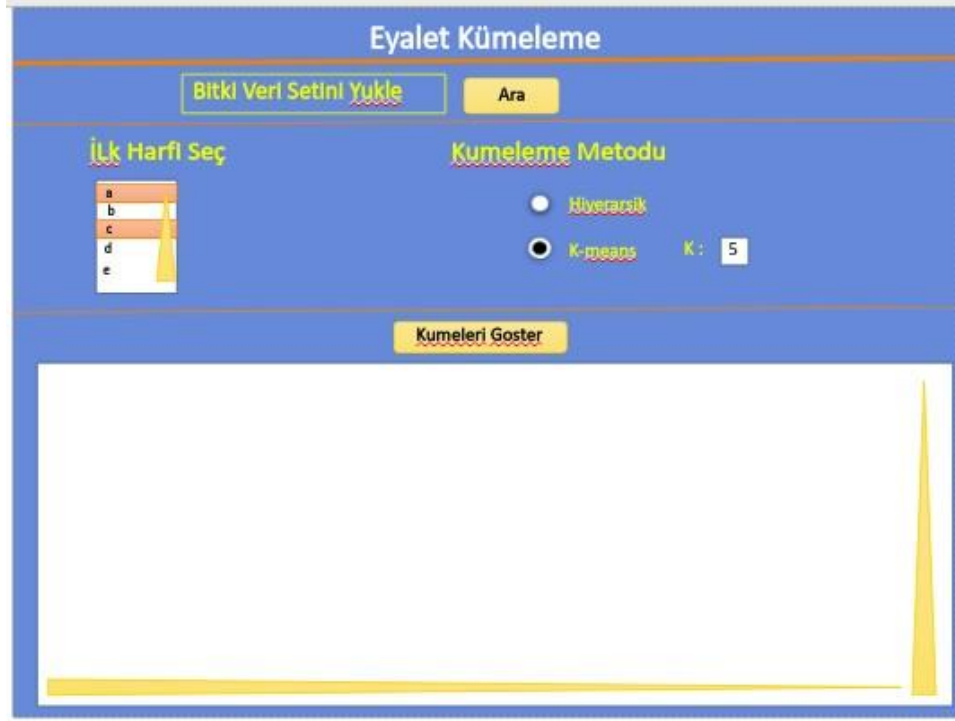
**abies**, eyalet\_kisaltması1, eyalet\_kisaltması2, s eyalet\_kisaltması3, ....  
**abies balsamea**, eyalet\_kisaltması4, eyalet\_kisaltması5, eyalet\_kisaltması6, ....  
**abies balsamea var. balsamea**, eyalet\_kisaltması7, eyalet\_kisaltması8, eyalet\_kisaltması9, ....  
**abies balsamea var. phanerolepi**, eyalet\_kisaltması10, eyalet\_kisaltması11, eyalet\_kisaltması12,  
.....

Programınızda aynı bitkinin tüm bu varyasyonlarına ait satırların, ilk kelimeyle tanımlanan bitkiye ait ek bilgiler olduğu kabul edilmelidir. Her bir satırda verilen eyaletlerin tümünün o bitkinin yetiştiği eyaletler olarak kayda geçirilmesi gerekecektir. Aynı bitki için verilen bilgilerde, aynı ya da farklı satırlarda, eyalet kısaltması tekrarları olabilir. Yani örneğin yukarıdaki satırlarda eyalet\_kisaltması4 ve eyalet\_kisaltması11 aynı kısaltmalar olabilir. Verilerin toplanmasıyla alakalı olan bu durum, o bitki türünün o eyalette yetişmekte olduğu bilgisine ek başka bir anlam etmemektedir.

Eyalet kısaltmalarının hangi eyalet isimlerine denk geldiği bilgisiye ayrı bir text dosyada verilmiştir. Bu dosyanın kullanıcı tarafından yüklenmesine gerek yoktur; programınız çalıştığında eyalet kısaltması ve eyalet ismi arasında eşleşme yapan bir varsayılan dosya olarak yüklenebilir. İsmi “stateabbr.txt” olan bu dosyanın formatı da aşağıdaki gibidir:

*eyalet\_kisaltması1 EyaletIsmi1*  
*eyalet\_kisaltması2 EyaletIsmi2*  
.....

1. Programınızın aşağıdaki şekilde verilene benzer bir grafik kullanım ara yüzü (GUI) olmalıdır. Detaylar ilerleyen paragraflarda anlatılacaktır.



Şekil 1

- Kullanıcı ilk olarak, bir örneği projeye birlikte verilen bitki veri seti dosyasını yükleyerek işlemlerine başlamalıdır (*Bitki Veri Setini Yükle*).
- Sonrasında, kullanıcının gruplama/kümeleme işlemini hangi harf ya da harflerle başlayan bitkilere göre yapabileceğini belirleyebileceği bir seçim yapma şansı olmalıdır (*İlk Harfi Seç*). Veri setinde çok sayıda bitki olduğu için, kullanıcıya böyle bir imkan vermek, O'nun farklı açılardan küme ve gruplar oluşturabilmesini sağlaması açısından programınızın kullanılabilirliğini arttıracaktır. Buradaki seçenekler 'a', 'b', 'c', ....., 'z' şeklinde olmalı ve ayrıca kullanıcı birden fazla harfi seçerek gruplama işlemini seçtiği harflerin tamamıyla başlayan bitkilere göre yapabilmelidir. Bunu çoklu seçme yapmaya imkan veren **Listbox** türü widget ile gerçekleyebilirsiniz. Widget'ınızı programınız ilk ayağa kalktığında ya da veri seti dosyası yüklendikten sonra harflerle doldurabilirsiniz.
- Kümelemeye esas olması için derslerde gördüğümüz "blogdata.txt" dosyasına benzer bir matris dosyası yaratmanız gerekecektir. Bu matris dosyasının satırları eyalet kısaltmalarıyla (ya da eyalet isimleriyle) başlamalı, kolonlardaysa bitki adları olmalıdır. Matrisin hücrelerindeyse, eğer o bitki o eyalette yetişiyorsa ilgili hücrede 1 değeri, yetişmiyorsa 0 değeri bulunmalıdır. Matrisinizi kümeleme işleminde kullanmak üzere "statePlantData.txt" isimli bir dosyada saklayabilirsiniz.

- Sonrasında kullanıcınız kümeleme işlemini hangi tekniğe göre (Hiyerarşik ya da K-Means) yapacağını seçecektir. K-means tercih etmesi durumunda, kaç adet “centroid” ile kümeleme yapılacağını da belirtebilecektir. Bitkilerin eyaletlere aitliği bir doğası gereği bir küme yapısı gösterdiğinden dolayı (bir bitki bir eyalette ya yetişiyordur ya da yetişmiyordur), Tanimoto benzerlik ölçütünün kullanılması daha doğru olacaktır. Tanimoto benzerlik ölçütünün implementasyonunu 4. laboratuvar alıştırmasında, filmler ve eleştirmenler bağlamında görmüştük. Aynı ölçütün kümeleme bağlamında benzerlik hesabı yapan fonksiyonu da “**clusters.py**” modülü içinde mevcuttur.
- Kullanıcı “Kümeleri Göster” butonuna tıkladığında kümeleme işleminin sonucunu en alttaki bölgede görebilmelidir. Hiyerarşik kümeleme için bu çıktı, Şekil 2’de gösterildiği gibi bir Dendrogram şeklinde olmalıdır. Bu çıktıyı elde edebilmek için **clusters.py** modülündeki **printclust** fonksiyonunu ele alıp, bu fonksiyonu çıktısını ara yüzünüzdeki “text widget”a bastırarak şekilde revize etmelisiniz. Bu revizyonu **clusters.py** modülü içinde değil, kendi projenize ekleyeceğiniz bir fonksiyon veya kod bloğu ile gerçekleştirmelisiniz. Eğer kullanıcınız K-Means kümeleme tekniğini seçecek olursa, çıktınız Şekil 3’deki gibi, oluşan küme üyelerinin listelenmesi şeklinde olmalıdır.

Text alanınızın hem dikey hem de yatay “Scrollbar”ları olmalı ve ayrıca oluşan çıktılarda, eyaletler kısaltmalarıyla değil, tam adlarıyla yer almalıdırlar.



Şekil 2

- Kullanıcınız bir kez kümeleme işlemini gerçekleştirdikten sonra, kullandığı kümeleme tekniği, başlangıç harfleri ve/veya “centroid” sayısını değiştirerek (K-Means kümeleme için) kümeleri baştan oluşturup gözlemleyebilmeli ve bu işlemleri dilediği kadar tekrarlayabilmelidir. Yeni bir kümeleme işleminin sonuçları bastırılmadan önce bir önceki işlemin sonuçları çıktı alanından temizlenmelidir. Kullanıcının her bir kümeleme girişimi için arka planda farklı bir matris

dosyası oluşturulmamalı, tek bir matris dosyasının içeriği sürekli güncellenerek kümelemede kullanılmalıdır.

Eyalet Kümeleme

Bitki Veri Setini Yükle Ara

İlk Harfi Seç Kumeleme Metodu

a b c d e

☐ Hierarşik ☒ K-means K: 5

Kumeleri Göster

Cluster 1 : {Arizona, Oklahoma, Utah }  
Cluster 2 : {Minnesota, Texas, California }  
Cluster 3 : {Manitoba, Rhode Island}  
Cluster 4 : {Ohio, Alberta, New Jersey }  
Cluster 5 : {Michigan}

Şekil 3

### Dikkat edilmesi gereken konular:

- Projenizi kodlamaya başlamadan önce problemi çözmek için hangi veri yapılarını kullanmanız gerektiğine karar verin.
- Mini projeler en fazla iki kişilik gruplar halinde yapılabilir. Diğer gruplarla fikir alışverişinde bulunabilirsiniz ancak kesinlikle kod paylaşımı yapmamanız gerekmektedir.
- Eğer bir web sayfası ya da bir kitaptan bir kod parçasını direkt alıp kullandıysanız, programınızın o bölgesine yazacağınız yorum satırları içinde bu durumu (programınızın o parçasını nereden alıp kullandığınızı) açıkça belirtin.
- Projenizi tek bir Python dosyası olarak teslim etmeniz beklenmektedir. Son teslim tarihi 28 Nisan Salı günü saat 14:00'dir. Bu tarihten sonra teslim edilen ödevler kabul edilmeyecektir.
- Derslerde üzerinden geçtiğiniz **clusters.py** modülünü projenizde "import" ederek kullanabilirsiniz, göndereceğiniz ödevlere bu modülü eklemeyiniz.

Eğer tek başınıza çalıştıysanız oluşturduğunuz Python dosya adı şu şekilde olmalıdır (tek bir dosya teslim edeceğiniz için ".zip" ya da ".rar" lamanıza gerek bulunmamaktadır):

*ad-soyad\_okulnumarası\_mp3.py*

Eğer iki kişilik gruplar halinde çalıştıysanız dosya adınız şu şekilde olmalı ve grup halinde tek bir ödev gönderimi yapmalısınız:

*ad\_soyad1\_numara1\_ad\_soyad2\_numara2\_mp3.py*

Ödev dosyalarınızı yukarıdaki isimlerle sakladıktan sonra Piazza sistemine yükleyiniz.

Dosya adlarında Türkçe karakterler (ı, ü, ğ, ö, ç, ş) ve büyük harfleri kullanmayınız.

### Değerlendirme Kriterleri:

Kod yapısı				İşlevsellik			
Class yapısı kullanımı (10)	Yorum satırları (5)	Anlamlı değişken isimleri (5)	Program hata vermeden yorumlanıyor ve çalışabiliyor mu? (10)	Ara yüz (GUI) görsel tasarım (11)	“Kümeleri Göster” butonu işlevi doğru çalışıyor mu (Hiyerarşik)? (13)	“Kümeleri Göster” butonu işlevi doğru çalışıyor mu (K-Means)? (13)	Kullanıcı ayarları değiştirip yeni baştan küme-leme yapabiliyor mu? (13)

Kolay gelsin.