

DATA STRUCTURES

Proje

CLASSIFICATION USING K-NEAREST NEIGHBORS (KNN) ALGORITHM

Ad-soyad, Numara

Emel Kayacı, 05180000087

İÇİNDEKİLER

1. Gerçekleştirilen Platform ve Dil ve Sürüm Adı	3
2. Programın kısa tanımı.....	3
3. Program için kullanılan bileşik veri tipleri, sınıfların ve metotların açıklamaları....	3
4. Örnek sonuçlar.....	5
5. Yazılım geliştirme için harcanan süreler.....	10
Ek 1: Program 1.2 a seçeneği: Sözdekod / Algoritma / Yöntem (Mantık) Anlatımı...	10

1. Gerçekleştirilen Platform ve Dil ve Sürüm Adı

Proje Netbeans 8.2 platformunda Java 8 ile gerçekleştirilmiştir.

2. Programın kısa tanımı

Bizden zambak bitkisinin çeşitli özelliklerini girdi olarak alıp üç farklı türünden hangisine ait olduğunu bulan algoritmanın tasarlanması isteniyor. Tür bulma işlemi kNN algoritması ile yapıp sınıflandırma işleminde karşılaştırılma için kullanılan veriler programa ilişkin iris.txt dosyasından kullanılmıştır.

Başarı ölçümü ile algoritmanın ne kadar başarılı çalıştığına dair başarı oranı bulunmuştur. Bu oran hesaplanırken sadece iris.txt dosyasında veriler kullanılmış olup ekleme veya silme işlemleri sonucu eklenen veya silinen veriler bu oranı etkilemeyecek şekilde düzenlenmiştir.

Klavyeden tüm özellikleri kullanıcı tarafından girilen bitkiler verisetine eklenmiş, sadece indisi verilen veriyi ve tüm verileri silen metotlar yazılmıştır. Ayrıca tüm verisetini listeleyen metot sayesinde verisetinde bulunan tüm bitkilerin özellikleri görüntülenebilmektedir.

3. Program için kullanılan bileşik veri tipleri, sınıfların ve metotların açıklamaları

- **Program için kullanılan bileşik veri tipleri ve açıklamaları:**
 - ***ArrayList<Zambak> zambakList:*** Zambak nesnesi tipinde verileri tutan ArrayList veri yapısıdır.
 - ***double[][] uzakliklar:*** İki boyutlu double tipinde matristir. Matrisin satır sayısı karşılaştırılan bitki sayısı kadar olup iki sütunu bulunur. İlk sütunda karşılaştırılan bitki örneğinin veri setindeki her bir bitki örneğinden uzaklığı, ikinci sütunda ise en yakın k adet bitkinin özelliklerini daha sonrasında görüntüleyebilmek için ArrayList içerisinde tutulduğu indeks değerleri bulunur.
 - ***double[][] sıraliUzakliklar:*** Bir üstte tanımlanmış olan veri tipinin ilk sütunundaki değerlere göre sıralanmış halidir. Satır ve sütunların ifade ettiği özellikler (sütunlar için sırasıyla uzaklık ve tutulan indeks değeri, satırlar için karşılaştırılan bitki sayısı) tamamen aynıdır.
 - ***ArrayList<Zambak> test ve ArrayList<Zambak> karsilastirma:*** başarıÖlç metodunda kullanılmak üzere test edilen ve karşılaştırılan zambakları tutan ArratList'lerdir.

- **Program için kullanılan metot ve açıklamaları:**

Not: Zambak sınıfı içerisinde kullanılan metotlar Zambak sınıfının açıklamasında belirtilmiştir.

- ***dosyaOku metodu:*** iris.txt dosyasında veriler *StringTokenizer* metodu ile her bir virgül ayıraç olacak şekilde token denilen parçalara ayrılmıştır. Ayrıştırılan bu veriler uygun özelliklere (çanak ve taç yaprakların uzunluk, genişlikleri ve zambak türü), uygun veri tipinde (yaprak uzunluk ve genişlikleri için double, tür için String) aktarılmıştır. Her bir çiçekten alınan bu veriler Zambak isminde birer nesnede tutulmuştur ve zambakların tutulduğu listeye aktarılmıştır.
- ***veriEkle metodu:*** Klavyeden listedeki verisetine yeni örnek çiçek verisi (öznitelikler ve sınıf) ekler.
- ***veriSil metodu:*** İndisi verilen veriyi verisetinden siler.
- ***tümünüSil metodu:*** Tüm verileri siler.

- **listeyiYazdir metodu:** Verisetindeki tüm değerleri zambak listesini forEach döngüsü ile dolaşarak görüntüler.
- **bitkiyiSınıflandır metodu:** kNN algoritması için k değeri ile sınıflandırılacak çiçek özelliklerini alır. Sınıflandırmak için alınan her bir çiçeği Zambak nesnesinde tutar.
 - **uzaklikBul metodu** ile özellikleri alınan çiçeği verisetindeki tüm kayıtlarla karşılaştırır.
 - **uzakliklariSiralama metodu:** Uzaklıkların tutulduğu iki boyutlu matris *sort* metodu kullanılarak küçükten büyüğe sıralanmıştır.
 - **türüTahminEt metodu:** *uzaklikBul* metodunun döndürdüğü matrisin sıralanmış hali kullanılarak ilk k tane çiçeğin türlerine bakarak en çok sayıda hangi türden çiçek varsa türünü bilmediğimiz çiçek bu türden sayılmıştır.
 - **enYakinZambaklariYazdir metodu:** En yakın k adet bitkinin özellikleri, uzaklıkları ve hangi sınıflardan oldukları ekrana listelenmiştir.
 - **bitkiyiSınıflandır metodu** içerisinde kullanılan bu metotlar dışında ayrıca en yakın zambakların çeşitli özellikleri ekrana yazdırıldıktan sonra tahmin edilen tür de ekrana yazdırılmıştır.
- **yaprakVerisiAl metodu:** Kullanıcıdan hem *veriEkle* metodunda hem de *bitkiyiSınıflandır* metodunda bitki verileri istendiğinden bitki verilerini kullanıcıdan alan bir metot oluşturulmuştur.
- **başariÖlç metodu:** Test ve karşılaştırma verileri için ArrayListler oluşturulmuş olup her bir bitki örneğinin sonunda yer alan 10'ar veri test verisi, geri kalan 120 veri ise karşılaştırma verisi olarak ayrılmıştır. Doğru tahmin edilen bitki sayısının bulunması için *testVeriSınıflandır* metodu kullanılmıştır.
 - **testVeriSınıflandır metodu:** Test verilerinin herbirini *uzaklikBul*, *uzakliklariSiralama*, *türüTahminEt* metodlarını kullanarak kalan 120 veri üzerinden sınıflandırmıştır. Tahmin edilen tür ile gerçek tür *equals* metodu ile karşılaştırılmış ve doğru tahmin sayısı bulunmuştur.
 - **başariOraniÖlç metodu:** Başarı oranı doğru sınıflandırılan bitki sayısının verisetinde test amaçlı kullanılan toplam bitki sayısına (30) bölünmesiyle hesaplanmış ve ekrana yazdırılmıştır.
- **Program için kullanılan sınıf ve açıklamaları:**
 - **Sınıflandırma sınıfı:** 1.2 a, 1.2 b, 1.2 c, 1.2 d ve e hesaplamalarını yapan metotları içeren sınıftır. Ayrıca main metodunu içerir.
 - **Zambak sınıfı:** Zambak bilgilerini içeren değişkenleri bulundurmaz. Zambakların taç ve çanak uzunluk ve genişliklerini temsil eden "double" tipinde değişkenleri ve "String" tipinde tür değişkenini bulundurmaz. Zambak tipinden nesneye özelliklerini atayan ve değişken parametrelere sahip constructorlar bulunur. Değişkenler private olarak tanımlandığından dışarıdan erişim olabilmesi için get ve set metotları bulunur. Sınıftan oluşturulan bir zambak objesinin özelliklerini yazdıran bir toString metodu vardır.

4. Örnek sonuçlar

Program menüsü:

```
MENÜ
Bitki sınıflandırması yapmak için 1'i
Verisetine yeni örnek çiçek verisi eklemek için 2'yi
Verisetinden çiçek verisi silmek için 3'ü
Başarı ölçümü için 4'ü
Verisetindeki tüm çiçek verilerini görmek için 5'i tıklayınız.
Menüden seçiminizi yapınız.
```

b seçeneği (bitki sınıflandırması) için istenen liste örneği:

```
Menüden seçiminizi yapınız.
1
kNN algoritmasında kullanılacak k değerini giriniz: 2

Aşağıdaki seçeneklere göre seçiminizi giriniz. (1/2/3)
1. Sınıflandırmak istenilen çiçeğin 4 adet özelliği girdi olarak alınacak.
2. Sınıflandırmak istenilen çiçeğin 2 adet çanak yaprak özelliği (uzunluk ve genişlik) girdi olarak alınacak.
3. Sınıflandırmak istenilen çiçeğin 2 adet taç yaprak özelliği (uzunluk ve genişlik) girdi olarak alınacak.
1
Çanak yaprak uzunluğunu giriniz: 2.3
Çanak yaprak genişliğini giriniz: 3.4
Taç yaprak uzunluğunu giriniz: 5.6
Taç yaprak genişliğini giriniz: 4.2
Örnek  ÇanUz  ÇanGen  TaçUz  TaçGen  Tür  Uzaklık
1      4.9    2.5     4.5    1.7    Iris-virginica  3.876854394996026
2      5.8    2.8     5.1    2.4    Iris-virginica  4.012480529547776
Tahmin edilen tür: Iris-virginica
```

c seçeneği (başarı ölçümü) için istenen liste örneği:

Iris-setosa türü için test verileri:

```
Menüden seçiminizi yapınız.
4
kNN algoritmasında kullanılacak k değerini giriniz: 1
Örnek  ÇanUz  ÇanGen  TaçUz  TaçGen  Tür  Uzaklık
1      5.1    3.5     1.4    0.3    Iris-setosa  0.14142135623730917
Tahmin edilen tür: Iris-setosa
Gerçek tür: Iris-setosa

Örnek  ÇanUz  ÇanGen  TaçUz  TaçGen  Tür  Uzaklık
1      4.4    2.9     1.4    0.2    Iris-setosa  0.6244997998398398
Tahmin edilen tür: Iris-setosa
Gerçek tür: Iris-setosa

Örnek  ÇanUz  ÇanGen  TaçUz  TaçGen  Tür  Uzaklık
1      4.4    3.0     1.3    0.2    Iris-setosa  0.20000000000000018
Tahmin edilen tür: Iris-setosa
Gerçek tür: Iris-setosa
```

Örnek 1	ÇanUz 5.0	ÇanGen 3.4	TaçUz 1.6	TaçGen 0.4	Tür Iris-setosa	Uzaklık 0.22360679774997896
Tahmin edilen tür: Iris-setosa						
Gerçek tür: Iris-setosa						
Örnek 1	ÇanUz 5.4	ÇanGen 3.9	TaçUz 1.7	TaçGen 0.4	Tür Iris-setosa	Uzaklık 0.3741657386773947
Tahmin edilen tür: Iris-setosa						
Gerçek tür: Iris-setosa						
Örnek 1	ÇanUz 4.9	ÇanGen 3.0	TaçUz 1.4	TaçGen 0.2	Tür Iris-setosa	Uzaklık 0.14142135623730986
Tahmin edilen tür: Iris-setosa						
Gerçek tür: Iris-setosa						
Örnek 1	ÇanUz 5.1	ÇanGen 3.8	TaçUz 1.5	TaçGen 0.3	Tür Iris-setosa	Uzaklık 0.14142135623730953
Tahmin edilen tür: Iris-setosa						
Gerçek tür: Iris-setosa						
Örnek 1	ÇanUz 4.6	ÇanGen 3.1	TaçUz 1.5	TaçGen 0.2	Tür Iris-setosa	Uzaklık 0.14142135623730964
Tahmin edilen tür: Iris-setosa						
Gerçek tür: Iris-setosa						
Örnek 1	ÇanUz 5.4	ÇanGen 3.7	TaçUz 1.5	TaçGen 0.2	Tür Iris-setosa	Uzaklık 0.10000000000000053
Tahmin edilen tür: Iris-setosa						
Gerçek tür: Iris-setosa						
Örnek 1	ÇanUz 5.0	ÇanGen 3.4	TaçUz 1.5	TaçGen 0.2	Tür Iris-setosa	Uzaklık 0.14142135623730964
Tahmin edilen tür: Iris-setosa						
Gerçek tür: Iris-setosa						

Iris-versicolor türü için test verileri:

Örnek 1	ÇanUz 5.7	ÇanGen 2.8	TaçUz 4.5	TaçGen 1.3	Tür Iris-versicolor	Uzaklık 0.3162277660168378
Tahmin edilen tür: Iris-versicolor						
Gerçek tür: Iris-versicolor						
Örnek 1	ÇanUz 6.1	ÇanGen 2.9	TaçUz 4.7	TaçGen 1.4	Tür Iris-versicolor	Uzaklık 0.14142135623730995
Tahmin edilen tür: Iris-versicolor						
Gerçek tür: Iris-versicolor						
Örnek 1	ÇanUz 5.8	ÇanGen 2.7	TaçUz 3.9	TaçGen 1.2	Tür Iris-versicolor	Uzaklık 0.14142135623730964
Tahmin edilen tür: Iris-versicolor						
Gerçek tür: Iris-versicolor						
Örnek 1	ÇanUz 4.9	ÇanGen 2.4	TaçUz 3.3	TaçGen 1.0	Tür Iris-versicolor	Uzaklık 0.1414213562373093
Tahmin edilen tür: Iris-versicolor						
Gerçek tür: Iris-versicolor						

Örnek	ÇanUz	ÇanGen	TaçUz	TaçGen	Tür	Uzaklık
1	5.5	2.5	4.0	1.3	Iris-versicolor	0.300000000000000016
Tahmin edilen tür: Iris-versicolor						
Gerçek tür: Iris-versicolor						
Örnek	ÇanUz	ÇanGen	TaçUz	TaçGen	Tür	Uzaklık
1	5.6	3.0	4.1	1.3	Iris-versicolor	0.1732050807568884
Tahmin edilen tür: Iris-versicolor						
Gerçek tür: Iris-versicolor						
Örnek	ÇanUz	ÇanGen	TaçUz	TaçGen	Tür	Uzaklık
1	5.6	3.0	4.1	1.3	Iris-versicolor	0.1732050807568884
Tahmin edilen tür: Iris-versicolor						
Gerçek tür: Iris-versicolor						
Örnek	ÇanUz	ÇanGen	TaçUz	TaçGen	Tür	Uzaklık
1	6.4	2.9	4.3	1.3	Iris-versicolor	0.200000000000000018
Tahmin edilen tür: Iris-versicolor						
Gerçek tür: Iris-versicolor						
Örnek	ÇanUz	ÇanGen	TaçUz	TaçGen	Tür	Uzaklık
1	4.9	2.4	3.3	1.0	Iris-versicolor	0.3872983346207412
Tahmin edilen tür: Iris-versicolor						
Gerçek tür: Iris-versicolor						
Örnek	ÇanUz	ÇanGen	TaçUz	TaçGen	Tür	Uzaklık
1	5.6	3.0	4.1	1.3	Iris-versicolor	0.22360679774997935
Tahmin edilen tür: Iris-versicolor						
Gerçek tür: Iris-versicolor						

Iris-virginica türü için test verileri:

Örnek	ÇanUz	ÇanGen	TaçUz	TaçGen	Tür	Uzaklık
1	6.9	3.2	5.7	2.3	Iris-virginica	0.26457513110645947
Tahmin edilen tür: Iris-virginica						
Gerçek tür: Iris-virginica						
Örnek	ÇanUz	ÇanGen	TaçUz	TaçGen	Tür	Uzaklık
1	6.9	3.1	5.4	2.1	Iris-virginica	0.3605551275463994
Tahmin edilen tür: Iris-virginica						
Gerçek tür: Iris-virginica						
Örnek	ÇanUz	ÇanGen	TaçUz	TaçGen	Tür	Uzaklık
1	5.8	2.7	5.1	1.9	Iris-virginica	0.0
Tahmin edilen tür: Iris-virginica						
Gerçek tür: Iris-virginica						
Örnek	ÇanUz	ÇanGen	TaçUz	TaçGen	Tür	Uzaklık
1	6.9	3.2	5.7	2.3	Iris-virginica	0.22360679774997935
Tahmin edilen tür: Iris-virginica						
Gerçek tür: Iris-virginica						

```

Örnek  ÇanUz  ÇanGen  TaçUz  TaçGen  Tür  Uzaklık
1       6.9    3.2    5.7    2.3    Iris-virginica  0.300000000000000016
Tahmin edilen tür: Iris-virginica
Gerçek tür: Iris-virginica

Örnek  ÇanUz  ÇanGen  TaçUz  TaçGen  Tür  Uzaklık
1       6.9    3.1    5.4    2.1    Iris-virginica  0.360555127546399
Tahmin edilen tür: Iris-virginica
Gerçek tür: Iris-virginica

Örnek  ÇanUz  ÇanGen  TaçUz  TaçGen  Tür  Uzaklık
1       6.3    2.7    4.9    1.8    Iris-virginica  0.24494897427831777
Tahmin edilen tür: Iris-virginica
Gerçek tür: Iris-virginica

Örnek  ÇanUz  ÇanGen  TaçUz  TaçGen  Tür  Uzaklık
1       6.5    3.2    5.1    2.0    Iris-virginica  0.22360679774997935
Tahmin edilen tür: Iris-virginica
Gerçek tür: Iris-virginica

Örnek  ÇanUz  ÇanGen  TaçUz  TaçGen  Tür  Uzaklık
1       6.3    3.4    5.6    2.4    Iris-virginica  0.2449489742783171
Tahmin edilen tür: Iris-virginica
Gerçek tür: Iris-virginica

Örnek  ÇanUz  ÇanGen  TaçUz  TaçGen  Tür  Uzaklık
1       6.1    3.0    4.9    1.8    Iris-virginica  0.282842712474618
Tahmin edilen tür: Iris-virginica
Gerçek tür: Iris-virginica

Başarı oranı: 1.0

```

d seçeneği (ekleme ve silme işlemleri) için istenen liste örneği:

Veri eklenmesi:

```

Menüden seçiminizi yapınız.
2
Çanak yaprak uzunluğunu giriniz: 1.2
Çanak yaprak genişliğini giriniz: 1.2
Taç yaprak uzunluğunu giriniz: 1.2
Taç yaprak genişliğini giriniz: 1.2
Tür giriniz: Iris-versicolor
Programdan çıkmak için e dışındaki herhangi bir tuşa basınız.
e

```


Verinin eklendiğinin kontrolü olarak tüm verilerin yazdırılması:

```
Çanak yaprak uzunluğu: 6.2
Çanak yaprak genişliği: 3.4
Taç yaprak uzunluğu: 5.4
Taç yaprak genişliği: 2.3
Tür:Iris-virginica

Çanak yaprak uzunluğu: 5.9
Çanak yaprak genişliği: 3.0
Taç yaprak uzunluğu: 5.1
Taç yaprak genişliği: 1.8
Tür:Iris-virginica

Çanak yaprak uzunluğu: 1.2
Çanak yaprak genişliği: 1.2
Taç yaprak uzunluğu: 1.2
Taç yaprak genişliği: 1.2
Tür:Iris-versicolor
```

Eklediğimiz veri verisetinin sonuna eklendi.

Programdan çıkmak için e dışındaki herhangi bir tuşa basınız.

Tüm verilerin silinmesi:

```
Menüden seçiminizi yapınız.
3
İndisi girilen veriyi silmek için 1'i
Tüm verileri silmek için 2'yi tıklayınız.
2
Listede bulunan tüm veriler silindi.
Programdan çıkmak için e dışındaki herhangi bir tuşa basınız.
e
Menüden seçiminizi yapınız.
5
Listede veri bulunmamaktadır.
```

Sadece indisi verilen verinin silinmesi:

```
Verisetindeki tüm çiçek verilerini görmek için 5'i tıklayınız.
Menüden seçiminizi yapınız.
5
Çanak yaprak uzunluğu: 5.1
Çanak yaprak genişliği: 3.5
Taç yaprak uzunluğu: 1.4
Taç yaprak genişliği: 0.2
Tür:Iris-setosa

Çanak yaprak uzunluğu: 4.9
Çanak yaprak genişliği: 3.0
Taç yaprak uzunluğu: 1.4
Taç yaprak genişliği: 0.2
Tür:Iris-setosa

Çanak yaprak uzunluğu: 4.7
Çanak yaprak genişliği: 3.2
Taç yaprak uzunluğu: 1.3
Taç yaprak genişliği: 0.2
Tür:Iris-setosa
```

0 indisinde bulunan veri

Verinin silindiğinin kontrolü olarak tüm verilerin yazdırılması:

```
Menüden seçiminizi yapınız.
3
İndisi girilen veriyi silmek için 1'i
Tüm verileri silmek için 2'yi tıklayınız.
1
Silinecek verinin indisini giriniz: 0
0 indisinde bulunan veri silindi.
Programdan çıkmak için e dışındaki herhangi bir tuşa basınız.
e
Menüden seçiminizi yapınız.
5
Çanak yaprak uzunluğu: 4.9
Çanak yaprak genişliği: 3.0
Taç yaprak uzunluğu: 1.4
Taç yaprak genişliği: 0.2
Tür:Iris-setosa
Eski indis: 1
Yeni indis: 0
Çanak yaprak uzunluğu: 4.7
Çanak yaprak genişliği: 3.2
Taç yaprak uzunluğu: 1.3
Taç yaprak genişliği: 0.2
Tür:Iris-setosa
Eski indis: 2
Yeni indis: 1
```

Not: Yukarıdaki ekran görüntüleri örnek amaçlı olduğu için sadece elemanları ilgilendiren indis/yerlerin yanlarındaki birkaç verinin ekran görüntüsü alınmıştır. (Örneklerde iris.txt dosyasındaki 150 veri kullanılmıştır.)

e seçeneği (listeleme) için istenen liste örneği:

Yukarıdaki örneklerde gösterilmiştir.

2.5 Yazılım geliştirme için harcanan süreler

- **Tasarım:** 2 gün
- **Gerçekleştirim:** 3-4 saat
- **Test:** 3 saat

Ek 1: Program 1.2 a seçeneği: Sözdekod / Algoritma / Yöntem (Mantık) Anlatımı

Elimizde türünü bilmediğimiz ancak bazı özelliklerini bildiğimiz bitki örneği bulunmaktadır. Bu çiçeğin hangi türe ait olduğunu 1.2 a seçeneğinde verilen k en yakın komşu algoritması ile bulabiliriz.

Bu algoritmada öncelikle k değeri belirlenir. Türünü araştırdığımız bitkinin diğer bitkilerden uzaklığı öklit uzaklığı ile hesaplanır. Bu uzaklıkları sıralar ve ilk k adet bitkiyi türünü aradığımız bitkiye en çok benzeyen bitkiler olarak ayırırız.

Benzer k adet bitkinin türlerine bakarız. Bu türler arasından sayıca en fazla olan tür aradığımız bitkinin türü olmuş olur. Eğer tür sayısı açısından iki veya daha fazla tür arasında eşitlik olursa aradığımız bitkiye en çok benzeyen bitkinin türünde (k = 1) sınıflandırma yaparız.