



Fenerbahçe Üniversitesi
BLM 103 – Algoritmalar ve Programlama I
Yapay Zeka ile Kanser Tespiti
Proje İçeriği
Veriliş Tarihi: 3.12.2019

Teslim Tarihi ve Yeri: 7.1.2020, Ders Saatlerinde, Ders Sınıfında ve Elektronik olarak

1. Tanım:

Bu proje yapay zeka uygulamalarında kullanılan bir algoritma olan kNN (k Nearest Neighborhood, En Yakın k Komşu) algoritması gerçekleştirilecektir. Gerçeklenen algoritma ile, UC Irvine Üniversitesi'nin sağlamış olduğu Göğüs Kanseri verileri işlenecektir. Hasta olup olmadığı belli olmayan bir kişinin verileri sisteme beslenerek, hastalık tahmini yapılacaktır.

2. Proje Ekibi:

Proje 1 veya 2 kişilik ekiplerden oluşacaktır. Her bir proje ekibinin bir sorumlusu olacaktır. Öğrenciler 1 veya 2 kişilik kendi proje ekiplerini ve proje sorumlusunu belirlemelidirler.

Ekiplerin kurulması ve proje sorumlusunun belirlenmesi en geç **10.12.2019** tarihine kadar tamamlanmalıdır. Ekip sorumluları, LMS'te açılmış olan "Proje Ekip Sorumluların Takımlarını Bildirmesi" başlığının altına, ekip üyelerinin isimlerini göndermelidirler.

LMS adresi: <http://levent.tc/lms/>

3. Kullanılacak Araçlar:

Proje kapsamında, ödev ve LAB'larda kullanılan Microsoft'un derleyicisi olan Visual Studio Community kullanılacaktır.

4. Tasarım Gereksinimleri

kNN algoritması ile, UC Irvine Üniversitesi veritabanından alınmış göğüs kanser verileri işlenerek, yeni bir kişinin verileri sisteme yüklenerek hasta olup olmadığı tahmin edilecektir.

Eğitim veriseti adresi: http://levent.tc/courses/blm103/projeler/proje1/breast-cancer-wisconsin_egitim.txt

Test veriseti adresi: http://levent.tc/courses/blm103/projeler/proje1/breast_cancer_wisconsin_test.txt

Veriseti dosyasını indirdiğinizde, her bir satırda virgüller ile ayrılmış 11 sayı görülecektir. Bu sayıların ifade ettiği anlamlar aşağıda verilmektedir.

Öğretim Elemanı: Dr. Vecdi Emre Levent, emre.levent@fbu.edu.tr, İzinsiz Kopyalanamaz

1. Hastadan alınan örneğin numarası
2. Clump Thickness
3. Uniformity of Cell Size
4. Uniformity of Cell Shape
5. Marginal Adhesion
6. Single Epithelial Cell Size
7. Bare Nuclei
8. Bland Chromatin
9. Normal Nucleoli
10. Mitoses
11. Sınıf Bilgisi (2 ise hasta değil, 4 ise hasta'yı ifade etmektedir)

Eğitim verisetinde 600 satır yani 600 örnek vardır. Projenin amacı mevcut hasta bilgilerini kullanarak, yeni bir bireyin değerlerini sisteme girerek hastalık tahmini yapmaktır. Bunun için kNN algoritması kullanılacaktır.

Test verisetinde ise 83 örnek vardır. Bu örnekler eğitim setine göre kNN algoritması uygulanarak test edileceklerdir.

kNN algoritması bir örneğin, veriseti içindeki kendisine uzaklık olarak en yakın olan k adet örneğin sınıf bilgisi en çok ne ise, sınıf bilgisi olarak bu değeri alır. Algoritmanın adımları aşağıda sıralanmaktadır.

1. En yakın k komşu sayısı için bir k değeri belirle, tek sayı olmalıdır. (3, 5, 7 ... gibi)
2. Test edilecek veriyi, verisetindeki tüm verilere göre uzaklığını bul
3. Uzaklıkları küçükten büyüğe doğru sırala
4. Sıralanmış örneklerin en küçük uzaklığa sahip k örneği al
5. K örneğin içindeki sınıf değeri en çok ne ise, test edilen verinin sınıf değeri olarak belirle

Sayısal Örnek

Elimizde 4 örnek bulunan bir kağıt mendilin kaliteli olup olmadığının ölçümü yapılmış bir veriseti vardır. Kağıtların kaliteli olup olmadığı, bu işin uzmanları tarafından belirlenmiştir. Veriseti aşağıdadır.

| X1 = Asit Dayanıklılığı (Saniye) | X2 = Esneme (Kg/m2) | Sınıflandırma |
|----------------------------------|---------------------|---------------|
| 7 | 7 | Kötü |
| 7 | 4 | Kötü |
| 3 | 4 | İyi |
| 1 | 4 | İyi |

Elimize yeni bir kağıt geldi. Bu kağıtın asit dayanıklılığı (X1) 3 ve esneme değeri (X2) 7 olarak geldi. Bu kağıtı bir uzmana göstermeden, mevcut verisetini ve kNN yapay zeka algoritmasını kullanarak, kağıtın kalitesini nasıl belirlenebilir?

1. İlk aşama k değerini belirlemektedir. K değeri 3 seçilmiştir (Örnek sayımız az olduğu için 3 seçildi, genellikle uygulamalarda 3 veya 5 seçilmektedir).
2. Eğitim verisi ile test edilen (3, 7) arasındaki uzaklıkları bul, uzaklık için Öklid yöntemi kullanılacaktır.

| X1 = Asit Dayanıklılığı (Saniye) | X2 = Esneme (Kg/m2) | Uzaklıklar |
|----------------------------------|---------------------|--------------------------|
| 7 | 7 | $(7-3)^2 + (7-7)^2 = 16$ |
| 7 | 4 | $(7-3)^2 + (4-7)^2 = 25$ |

| | | |
|---|---|--------------------------|
| 3 | 4 | $(3-3)^2 + (4-7)^2 = 9$ |
| 1 | 4 | $(1-3)^2 + (4-7)^2 = 13$ |

3. Uzaklığa göre en küçükten en büyüğe doğru sırala

| X1 = Asit Dayanıklılığı (Saniye) | X2 = Esneme (Kg/m2) | Uzaklıklar | Sınıflar |
|----------------------------------|---------------------|--------------------------|----------|
| 3 | 4 | $(3-3)^2 + (4-7)^2 = 9$ | İyi |
| 1 | 4 | $(1-3)^2 + (4-7)^2 = 13$ | İyi |
| 7 | 7 | $(7-3)^2 + (7-7)^2 = 16$ | Kötü |
| 7 | 4 | $(7-3)^2 + (4-7)^2 = 25$ | Kötü |

4. k örneği al, k sayısı 3 olduğu için ilk 3 örnek alınacaktır.

| X1 = Asit Dayanıklılığı (Saniye) | X2 = Esneme (Kg/m2) | Uzaklıklar | Sınıflar |
|----------------------------------|---------------------|--------------------------|----------|
| 3 | 4 | $(3-3)^2 + (4-7)^2 = 9$ | İyi |
| 1 | 4 | $(1-3)^2 + (4-7)^2 = 13$ | İyi |
| 7 | 7 | $(7-3)^2 + (7-7)^2 = 16$ | Kötü |

5. k örneğin içindeki en çok sınıf değeri ne ise, sonuç olarak belirle.

3 örneğin içerisinde, iyi, iyi ve kötü sınıf değerleri vardır. İyi değeri daha çok olduğu için, verilmiş test örneği iyi olarak tespit edilmiştir.

Gerçekleme Yöntemi:

Verilen eğitim ve test verilerini dosyadan okuyarak belleğe yükleyiniz.

Test verilerinin her bir satırını alıp, örnek numarası bölümünü haricinde sınıf bilgisine bakmadan diğer 9 özelliğine göre sınıfını bulun. Bulunan sınıf, test verisinde yazan sınıf (aslında olması gereken sınıf) aynı ise, doğru, farklı ise yanlış sınıflandırılmıştır. Tüm test verileri için bu işlemi yapıp kaç adet doğru kaç adet yanlış sınıflandırıldı. Yüzde olarak başarı nedir ekrana yazdırınız.

Ayrıca istendiğinde kullanıcıdan komut satırından 9 parametre alarak, hasta mı değil mi ekrana gösteren bir menü bulunmalıdır.

Programın ilk seçeneği test dosyasını hesapla

İkinci seçeneği ise kullanıcıdan parametre al

olmalıdır.

5. Notlandırma ve Proje Teslimi:

Bu başlık "Yapay Zeka ile Kanser Tespiti" projesinin teslimi ve notlandırılması hakkında bilgiler içermektedir.

5.1. Notlandırma:

Öğretim Elemanı: Dr. Vecdi Emre Levent, emre.levent@fbu.edu.tr, İzinsiz Kopyalanamaz

Projenin **iki** ana değerlendirme kriteri vardır. Her iki kriter 50 şer puandır.

İlk kriter “Yapay Zeka ile Kanser Tespiti” ‘nin **doğru çalıştırılmasıdır**.

İkinci kriter ise **Proje Teslim Dokümanı ve Sunumdur**.

- **Proje Teslim Dokümanı:**

Öğrenciler, proje raporlarını verilen “Proje Teslim Dokümanı” ‘nın içerisini doldurarak yapacaklardır.

Proje Teslim Dokümanı: http://levent.tc/courses/blm103/BLM103_proje_teslim_dokumani.doc

- **Proje Sunumu:**

Powerpoint üzerinde ortalama 5 dakika (4-6 dakika arası) sürecek bir sunum hazırlayarak kayıt etmelidirler. Kayıt işlemi, cep telefonu veya bilgisayar ekran kayıt yazılımları (Screen-Recorder, Bandicam vb...) ile yapılabilir.

Sunum, ekip üyeleri içinden biri tarafından, projenin nasıl yapıldığı, nasıl çalıştığı vb.. konularının powerpoint slaytları üzerinden anlatılırken kaydedilmesi ile olmalıdır. Sunum video’sunda powerpoint slaytları okunabilir ve konuşmacının sesinin anlaşılır olması gerekmektedir. Powerpoint slayt görünüm tasarımı istenildiği gibi yapılabilir.

Proje ekibinin tamamı, notlarını bu değerlendirmeye göre alırlar.

5.2. Teslim:

Projenin teslimi için aşağıdaki adımların gerçekleştirilmesi gerekmektedir. İstenen dosyaları sadece proje ekip sorumlusunun getirmesi, LMS ve Github (Çok yaygın bir açık kaynak kod paylaşım platformudur)’a yüklemelidir.

Proje Teslim Dokümanının, ders sınıfı ve saatinde, çıktılarının alınarak teslim edilmesi gerekmektedir.

Ayrıca LMS’te açılmış olan “Proje Teslim” sayfasına aşağıdaki dosyaların yüklenmesi gerekmektedir.

- Kaynak C kodu
- Hazırlanan powerpoint sunum dosyası (.ppt uzantılı dosya)
- Proje Teslim Dokümanı (Word formatında yüklenmelidir)
 - Dokümanın alt başlıkları doldurulmalıdır
 - Kaydedilen powerpoint sunum video’su youtube’a yüklenip, adresi, dokümanın sonuçlar bölümündeki açılmış yere link’i yazılmalıdır (Video’nun herkes’e görünür olmamasını istiyorsanız, youtube’a yükledikten sonra liste dışı seçeneğini seçerek, sadece link’e sahip olan kişilerin görmesini sağlayabilirsiniz).
 - LMS’e yüklenen tüm dosyalar (Kaynak C kodu, ppt uzantılı sunum dosyası ve Proje Teslim Dokümanını (PDF formatında)), github.com sitesine üye olup, yüklenip, dokümanın sonuçlar bölümündeki yere link’i yazılmalıdır.