

## YAZ 104 Temel Programlama II

Bahar 2020

Mini Proje 5

(80 puan)

26 Mayıs 2020

Bu mini projede, derslerde gördüğümüz doküman sınıflandırıcımızın Amerikan Şakalarını doğru bir şekilde sınıflandırmadaki performansını görme şansınız olacak. Yükleyeceğiniz bir veri seti üzerinden çeşitli türdeki Amerikan şakalarını içeriklerine bakarak kategorize etmeye çalışacaksınız. Veri setimiz sınıflandırıcınızı eğitmek üzere beş farklı kategoride şakalar içeriyor. Bu kategoriler “clinton”, “lawyer”, “math”, “medical”, ve “music”. Ayrıca veri setinde, sınıflandırıcınızın eğitiminde ve daha sonra test edilmesinde kullanmak üzere, her kategori için iki grup halinde şaka verileri mevcut. Eğitim veri setindeki şaka metinleriyle sınıflandırıcınızı eğittikten sonra, test veri setinden sınıflandırıcınızın daha önce görmemiş olduğu şakaların kategorilerini tahmin etmesini isteyeceksiniz.

Programınızın Şekil 1’de verilene benzer bir grafik kullanım ara yüzü (GUI) olmalıdır. Şekildeki renklendirmeler sadece bir örnektir, siz projenizde dilediğiniz renkleri kullanabilirsiniz. Ancak widget’ların yerleşimi şekildeki gibi veya ona çok yakın olmalıdır. **Dikkat!:** Projenizde Place Geometry Manager kullanmamalısınız; Pack ya da Grid Geometry Manager’den herhangi birini kullanabilirsiniz.

Amerikan Saka Sınıflandırıcısı

Eğitim Verisi Yükle

Eğitim Verisi Yüklendi...

Sınıflandırıcı

Naive-Bayes

0.3

Eşik Değer

clinton

lawyer

math

medical

music

Set

lawyer- 0.6

music- 0.2

clinton- 0.3

Seçiliyi Kaldır

Sınıflandırıcıyı Eğit

Eğitim Tamamlandı....

Sakayı buraya girin

Sınıflandırıcıyı Test Et

Girdiğiniz sakanın Kategorisi:

<Kategori>

Performans Testi

(.....)

Sınıflandırıcınızın Performans Test Sonucu

<%...>

Şekil 1

- Programınızda veri setinin uygun bir dizinde (örneğin programınızla aynı dizinde) bulunduğunu kabul edebilirsiniz. “Eğitim Verilerini Yükle” butonuna basıldığında, programınız veri setini ilgili dizinde yüklemeli ve yükleme işlemi sona erdiğinde de, butonun yanındaki durum panelinde “Eğitim Verileri Yüklendi” şeklinde bir ibareyle bu aşamanın sona erdiğini kullanıcıya bildirmelidir. Veri setinin tamamı yükleninceye kadar bu alanda herhangi bir şey yazmamalıdır.

- Daha sonra kullanıcı, dilediği takdirde, sınıflandırma aşamasında kullanmak üzere, dilediği kategoriler için eşik değerleri girebilmeli veya girdiği eşik değerlerini kaldırabilmelidir. Burada bir Listbox kullanabilir ve listbox'unuzu program çalıştığında içerdiği ilk eşik değerlerle doldurabilirsiniz. Kullanıcı combo-box'dan seçeceği herhangi bir kategori için, combo-box'un solundaki alana gireceği bir değerle o kategori için seçtiği eşik değeri belirleyebilmelidir. Benzer şekilde, kullanıcı daha önce bir kategori için set ettiği bir eşik değerden vazgeçmek istediğinde de kaldırmak istediği eşik değeri seçtikten sonra “**Seciliyi Kaldır**” butonuna tıklayarak bunu gerçekleştirebilmelidir (Belirlenmiş bir eşik değerinin kaldırılmasının ne anlama geldiğini düşünmelisiniz). Eğer kullanıcı eşik değeri olan bir kategori için tekrar bir eşik değer set etmek isterse, en son belirlediği eşik değeri geçerli olmalı ve Listbox'da bu değer görünmelidir.
- Kullanıcı bu aşamadan sonra “**Sınıflandırıcıyı Eğit**” butonuna basarak, yüklenmiş olan eğitim veri setiyle sınıflandırıcısını eğitecektir. Programınız eğitim aşaması tamamlandığında da bu durumu butonun yan tarafındaki durum panelinde “**Eğitim Tamamlanmıştır**” şeklinde bir ibareyle belirtmelidir. Eğitim aşaması tamamlanıncaya kadar bu alanda herhangi bir şey yazmamalıdır.
- Eğitim aşaması bittiğinde sıra sınıflandırıcının test edilmesine gelecektir. Kullanıcı ilgili Text alanına bir şakanın tam metnini yazıp ardından da “Sınıflandırıcıyı Test Et” et butonuna basarak girdiği şaka metninin kategorisinin ne olduğunu sınıflandırıcıdan tahmin etmesini isteyebilmelidir. Sonuç, **Kategori** yazılı alanda gösterilmelidir. Eğer kullanıcı herhangi bir metin girmeden sınıflandırma yapmaya kalkarsa, programınız kullanıcıyı uyarmalı ve bir şaka metni girmesini istemelidir.
- Test aşamasında, eğer elinizde tipini bildiğiniz Amerikan şaka metinleri varsa, dilediğiniz herhangi bir test şaka metnini kullanabileceğiniz gibi, veri setinde test amaçlı ayrılmış ve doğru cevapları belli olan şakaları da kullanabilirsiniz. Veri setindeki şaka metinlerini kullanarak sınıflandırıcınızın tahmin performansı hakkında da bir fikir sahibi olabilirsiniz.
- Son olarak kullanıcınız en alt bölümdeki “**Performans Testi**” butonuna basarak, sınıflandırıcının test veri setindeki bütün şakaları sınıflandırmadaki doğruluk oranını ölçüp hesaplamalı ve sonucu da ilgili alana (%’lik değer olarak) yazmalıdır. Performans testinin tamamlanması biraz zaman alabileceğinden, test devam ederken kullanıcının sabretmesini sağlamak adına, butonun yanındaki arka arkaya eklenen noktalar (.....) şeklinde testin sağlıklı bir şekilde ilerlemekte olduğu hissi verilmeli, test tamamlandığındaysa aynı alana “**Sınıflandırıcınızın Performans Test Sonucu**” ibaresi gelmelidir. Test sonuçlanıncaya kadar yüzdelik değerin gösterileceği alanda herhangi bir bilgi olmamalıdır.
- Kullanıcınız, eşik değerlerini değiştirerek ve/veya sınıflandırıcıyı tekrar tekrar eğiterek, yeni tahmin denemeleri ve performans testleri de yapabilmelidir. Her denemesinden önce eğitim setini baştan yüklemek zorunda olmaması gerekir.

- Eğitim ve test aşamalarında kullanacağınız veri setleri bu proje metniyle verilen “**data\_set.zip**” isimli dosyadadır. Bu sıkıştırılmış dosyayı açtığınızda göreceğiniz dizinlerden “learn” isimli olanı eğitim setini, “test” isimli olanı da test veri setini içermektedir. Her bir veri seti dizininde kategoriler için alt-dizinler ve bu alt-dizinlerde de her biri bir adet şaka metni içeren pek çok text dosya mevcuttur. Bu dosyaları uygun bir şekilde işleyip programınız içerisindeki veri yapılarına almanız gerekecektir.

### İşinize yarayacak bilgiler:

- Derslerde gördüğümüz **docclass** modülünü import edebilir ve içerisindeki çeşitli fonksiyonları kullanabilirsiniz. Projenizi gönderirken **docclass** modülünü göndermenize gerek yoktur.
- Kullandığınız Listbox widget’ların dikey Scrollbar’ları da olmalıdır.
- Açtığınız herhangi bir editörden Text Widget’iniz içerisine standart kopyalama (CTRL-Y) ve yapıştırma (CTRL-V) kısa yollarını kullanarak text aktarabilirsiniz.
- Text Widget içerisine yazılmış olan metinleri Text Widget’in **-get** metodunu kullanarak elde edebilirsiniz. Aşağıdaki bağlantılar size bir Text Widget’den okuma yapmak konusunda fikir verecektir:

<http://effbot.org/tkinterbook/text.htm>

<https://www.it-swarm.dev/tr/python/tkinter-metin-kutusu-widgetindan-giris-nasil-alinir/1071050763/>

- Kullanıcıyı uyarmak için MessageBox widget’ini kullanabilirsiniz. Aşağıdaki link’de MessageBox kullanımıyla ilgili basit örnekler bulabilirsiniz:

<https://docs.python.org/3.9/library/tkinter.messagebox.html>

### Dikkat edilmesi gereken konular:

- Projelerde bunun sonra Place Geometry Manager **KULLANMAMALISINIZ**; Pack ya da Grid Geometry Manager’den herhangi birini kullanabilirsiniz. Place Geometry Manager kullananlar projenin GUI kısmından puan alamayacaklardır.
- 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 26 Mayıs Salı günü saat 14:00’dir. Bu tarihten sonra teslim edilen ödevler kabul edilmeyecektir.

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\_mp5.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\_mp5.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 (5)	Yorum satırları (5)	Anlamlı değişken isimleri (5)	Program hata vermeden yorumlanıyor ve çalışabiliyor mu? (5)	Ara yüz (GUI) görsel tasarım (10)	“Girilen Şakanın Kategorisini Tahmin Etme ve Gösterme (15)	“Performans Testi yapılması ve sonucun alınması” (20)	“Parametrelerin değiştirilerek baştan sınıflandırma ve test yapılabilmesi” (15)

Kolay gelsin.