

## Yapay Zeka 2022/2

### Ödev No:1

**Ödev Konusu:** Genetik algoritma / Tepe tırmanma ile metin sınıflandırma

2 sınıflı en az 100 örnek içeren metin sınıflandırma veri kümesi bulunuz (olumlu / olumsuz, spor / ekonomi vb.).

Bireyler N adet kelime içeren bir listeden oluşacaklar. İlk kısmı 1. sınıfa, son kısım ise 2. sınıfa ait olmayı ifade edecektir.

Örnek bir birey ["güzel", "harika" ... "müthiş"] ["berbat", "kötü" ... "beğenmedim"] şeklinde olursa bu birey 100 örneği bu kelimeleri içermelerine göre sınıflandıracak ve bir sınıflandırma başarısı elde edecektir. Bu başarı değeri fitness fonksiyonu olarak kullanılacaktır.

Sınıflandırma metnin içerdiği kelimelerin kaçının 1. sınıfa, kaçının 2. sınıfa ait listede olduğu bulunarak bu değerlere göre işlem yapılabilir. Beraberlik durumunda sınıf rasgele belirlenebilir.

Başlangıçta kelimeler tüm bireylere rasgele seçilerek atanacak. Süreç boyunca **crossover**, **rasgele kelime değiştirme** gibi işlemler uygulanarak bireylerin fitness değeri yükseltilmeye çalışılacak.

Bireylerin kelime listeleri genetik algoritma ya da tepe tırmanma ile optimize edilecektir.

Denemelerde aşağıdaki hiper parametrelerden en az 2'si incelenecektir:

**N (bireylerin içerdiği listenin uzunluğu)**

Popülasyon büyüklüğü (birey sayısı)

**Mutasyon oranı (zaman içinde değişebilir)**

**Önemli Not: Deney sonuçlarınızı, bulgularınızı iyi ifade eden tablolar ve grafiklerle sununuz. Ör: jenerasyonlar ilerledikçe en iyi bireyin sınıflandırma başarısı ve bireylerin ortalama başarısı**

**Ödevin Son Teslim Tarihi:** 27 Nisan 2023 saat 08:59

**Ödevin Teslim Şekli:** [online.yildiz.edu.tr](https://online.yildiz.edu.tr)

Bu ödevde en çok 2 kişilik gruplar halinde çalışacaktır.

**Ödevde Teslim Edilecekler (Sisteme yükleyeceğiniz (sıkıştırılmış tek bir) dosyanın içeriği):**

1-Ödev Raporu (4/10 puan)

2-Programın çalıştırılabilir hali ve açıklama içeren kodları  
(değişkenlerin ne için kullanıldıkları, algoritmanın adımları) **(3/10 puan)**

3-Programınızın en az 3 çalışmasını içeren, maksimum 10 dakika  
uzunluğunda (gerekiyorsa hızlandırın) ve youtube'a yüklenmiş videonun adresi  
(video da 3 farklı veri kümesi için jenerasyon sayısı ile o jenerasyonun en  
iyi bireyinin sınıflandırma başarısı grafiği görülmelidir) **(3/10 puan)**

**Ödev Raporunun içeriği:**

- 5 farklı veri kümesi için üretilen en iyi kelime listeleri
- seçtiğiniz en az 2 hiperparametreye dair bulgularınızı açıklayan tablo ve grafikler
- Bulgularınıza dair yorumlarınız