

BİL3102 - Ödev

İnternet ortamında yazılmış Türkçe kayıtlardan oluşan Veri Kümesi için

Duygu (Sentiment) Analizi

Dersin Öğr. Elemanı: Yrd. Doç. Dr. Mete Eminağaoğlu

Ödev Konusu: Aşağıda çok kısa özet bilgiler verilmiştir. Detaylı açıklamalar, sadece bir kere, ders saatinde yapılacaktır. O derse gelemeyen, vb. öğrencilerin ödevi yanlış / eksik vb. anladığı için yetersiz bir ödev ilemesinden sadece o öğrencinin kendisi sorumludur.

Ödevin Son Teslim Tarihi: 11 Nisan 2018, 23:00 (TSİ)

Ödevin Sunumu: 13 Nisan 2018, 09:30 (TSİ)

DİKKAT! Ödevlerin teslimi yeterli değildir. 13 Nisan 2018 günü ders saatinde tüm öğrencilerin derse gelmesi ve ödevlerini sınıfta sunmaları zorunludur. Ödevini zamanında teslim etmiş olsa bile, 13 Nisan 2018 tarihindeki derse gelmeyen ve sunumu yapmayan öğrenciler de ödevden 0 (sıfır) alacaktır.

Ödevin Teslim Şekli:

CSC ÖBS (Moodle) sistemindeki ders sayfasında açılacak olan ödev yükleme (assignment) alanına; tüm dosyalar, rapor, vb. zip / rar sıkıştırılmış tek bir dosya olarak yüklenecektir.

Bu ödev tek kişiliktir. Her türlü yardım, vb. kopya / intihal olarak değerlendirilecektir ve yardım eden / yardım alan öğrencilerin hepsi bu ödevden 0 (sıfır) alacaktır.

Ödevin geç teslimi durumları:

1 gün geç iletilen ödevlerde; öğrenci o ödevden alacağı notun % 50'sini,
2 gün geç iletilen ödevlerde; öğrenci o ödevden alacağı notun % 25'ini,
3 gün geç iletilen ödevlerde; öğrenci o ödevden alacağı notun % 12.5'ini,
≥ 4 gün geç iletilen ödevlerde; öğrenci o ödevden alacağı notun % 1'ini
...alacaktır.

Örnek: Öğrenci, 100 üzerinden 80 alacağı düzeyde yaptığı ödevini;

1 gün geç iletirse; ödevinden 40 alacaktır	(100 üzerinden)
2 gün geç iletirse; ödevinden 20 alacaktır.	" "
3 gün geç iletirse; ödevinden 10 alacaktır.	" "
4, 5 vb. gün geç iletirse, ödevinden 0.8 alacaktır.	" "

Veri kümesinin içeriği:

"Odev-veriler.rar" adlı sıkıştırılmış dosya içerisinde, "metinler" klasörü ve bu klasör altında farklı uzunluklarda 135 adet kayıt bulunmaktadır ve her bir kayıt (yazı metni) ayrı bir txt dosyasındadır (1.txt, 2.txt, vb.). Bu kayıtlar, İnternet'teki çeşitli sosyal ağ, blog, forum sitelerinden alınmıştır. Duygu analizinde kategorik ayırım şu şekilde yapılacaktır:

- 1- Olumlu
- 2- Olumsuz
- 3- Belirsiz

Ödevde Yapılacaklar ve İstenenler:

- 1. **Aşama:** Bu veri kümesini kullanarak, derslerde size aktarılan bir veya birkaç yöntemi kullanarak 135 farklı kaydın her birisini 3 kategoriden birisine atayacaktır. Bu atamaları, ya program ekranında liste şeklinde görüntülenecek, ya da bilgisayarda bir dosya yaratıp oraya liste olarak yazdırılacaktır.

- Örnek:

21.txt	Olumsuz
22.txt	Belirsiz
...	
117.txt	Olumlu

- **2. Aşama:** Programın ekranından herhangi bir sözcük girildiğinde, bunun hangi olasılık değeri ile hangi duygu kategorisine atandığı hesaplanıp ekrana yazılacaktır.
 - *Örnek: ekrandan “kalite” sözcüğü girilir.*
Olumlu
p=0.0000342
 - *Örnek 2: ekrandan “tebeşir” sözcüğü girilir.*
Belirsiz
p=0.00000157
vb...
- Programınızda, hazır kütüphane / fonksiyon, vb kullanabilirsiniz, bu konuda bir kısıtlama yoktur.
- Kodlama kısmında, istediğiniz programlama dilini kullanabilirsiniz (C, C++, C#, .Net, Java, Python, Ruby, Swift, vb).

Ödevde Teslim Edilecekler:

1-Programın tüm kaynak kodları, bağlantılı kütüphane, dizinler, vb.

2-Kullanılan algoritmalar, vb. ile ilgili kısa bilgiler / notlar (istenirse kaynak kod içine de açıklamalar olarak eklenebilir).