

Doğal Dil İşlemeye Giriş

Ödev 1

1. Refik Anadol'un Büyük Dil Modeli Türkiye temalı konuşması ile Geoffrey Hinton'ın Yapay Zeka Atlası isimli röportajında ortak olan ve /veya olmayan anlatımlara göre dersimizle ilişkilendirilebilecek hangi bildirimleri yakalayabildiniz? Bu iki videodaki söyleşilerin, sizin DDİ ve Yapay Zeka alanında genel bilgi edinmenize, belki de bu alanda daha fazla gelişmek üzere araştırma yapmanıza katkı sağlayacak kavramlar bağlamında **anlatınız**.

Notlar: i) Refik Anadol videosu bağlamındaki araştırmalarınız ve yorumlarınız sadece DDİ dersinin içeriği ile ilgili olmak **zorundadır**.

ii) İki konuşmanın karşılaştırması **kesinlikle** düşünülemez. Geoffrey Hinton bilgisayar bilimine büyük katkıları olan yaşayan bilim insanı, Refik Anadol ise dünya çapında bir resim sanatçısıdır. Farklı disiplinlerden uygulama alanları olan dersimiz ve Yapay Zeka ile ilgili olarak aykırı kimliklerin **aynı olgulardan farklı düzeylerde** konuşulabildiği görülmektedir.

ii) Bu sorunun cevaplanması için her 2 videonun tamamen dinlenmesi zaman alacaktır. Bir dinletiden sonuç çıkarabilmek, düşüncelerini yazabilmek ve en önemlisi araştırma yapabilmek ancak dinlerken **not alarak** mümkün olacaktır.

iii) İngilizce röportajı anlayamadığınızı düşünüyorsanız açık olarak belirtmeniz gerekir. Mesleğinize özel, oldukça akıcı bir dile sahip röportaj olduğu için, asgari ölçüde anlaşılabilirlik sağlamanız **kariyeriniz bağlamında** beklenmektedir.

2. **Zemberek** (önceki ismi Kemik) <https://github.com/ahmetaa/zemberek-nlp> NLP tools for Turkish.

Python NLTK : Natural Language Tool Kit, dünya ölçeğinde çok kullanılan bir doğal dil işleme aracıdır.

Her iki DDİ aracı, dersimizde **metin işleme** ile ilgili gerçekleştirilen tüm süreçleri içermektedir. Uygulama dili olarak hangi geliştirme aracını çalışacağınız sizin tercihinizdir. Hangisi üzerinde daha detaylı araştırma yapmak istediğinize karar verdikten sonra, 4 hafta süresinde işlediğimiz **metin işleme** proseslerinden biri ya da birkaçına ilişkin **uygulama** yapmanız istenmektedir. Probleminiz için veri setine ihtiyacınız olduğunu biliyorsunuz. Hazır veri setlerini de araştırmanız önerilir.

Önemli Not: Bu sorunun cevaplanmasında, araştırma sürecinin ve çalışma aşamalarının anlatımı istenmektedir. Diğer bir ifade ile bir **kod ve ekran çıktısı** ödevin 2. sorusuna cevap olarak **değerlendirilmez**.

Değerlendirme: Sadece her iki soru üzerinde de çalışılmış olan ödevlerin değerlendirmeleri eşit ağırlıklıdır. Zira, her iki soru birbirlerinden farklı ve ayrı ölçüklere göre önemlidir.

Aynı ödevlerde **veren /alan araştırılmaz**; her ikisi de değerlendirme dışı kalır. Kanıt da istenmez ve kabul edilmez.