

# nlp-arge-odev

**Veri Seti :** <https://www.kaggle.com/datasets/rmisra/news-category-dataset>

 **NLP Ar-Ge Odaklı Ödev**

**Konu: Haber Başlıklarından Duygu Analizi ve Topic Modeling**

 **Amaç**

Haber başlığı verisi üzerinde doğal dil işleme adımlarını uygulayarak:

- Metni temizleme ve hazırlama (preprocessing)
- Metni sayısal hale getirme (vektörleştirme)
- Duygu analizi (sentiment analysis)
- Anlam kümesi çıkarımı (topic modeling)

yapmanız ve bunların performanslarını karşılaştırarak **deneysel gözlem yeteneğinizi**

geliştirmeniz beklenmektedir.

 **Veri Seti**

Aşağıdaki kaynaklardan biriyle çalışabilirsiniz:

- Kaggle: News Category Dataset
- BBC News başlıkları (manuel olarak toplanabilir)
- Kendi oluşturduğunuz Türkçe haber başlıkları listesi

 **Uygulamanız Gereken NLP Ön İşleme Adımları**

Her adımı uygulayıp neden o sırayla yaptığınızı açıklayın:

- Tokenization
- Lowercasing
- Stopword removal
- Lemmatization

- (İsteğe bağlı) POS tagging

## **Vektörleştirme Yöntemleri**

Aynı veriyi farklı yöntemlerle sayısallaştırın:

- CountVectorizer (BoW)
- TF-IDF

Her yöntemin avantaj/dezavantajlarını örnek çıktılarla gösterin.

## **Duygu Analizi (Sentiment Analysis)**

- İngilizce başlıklar için TextBlob ya da VADER
- Türkçe başlıklar için Zemberek, BERTurk ya da başka transformer tabanlı modeller

Pozitif / Negatif / Nötr gibi sınıflandırmalar yaparak kısa bir sonuç değerlendirmesi ekleyin.

## **Topic Modeling**

- LDA (Latent Dirichlet Allocation) ya da NMF yöntemlerini kullanarak başlıkları konu

kümelerine ayırın.

- Her topic'e anlamlı bir isim vermeye çalışın.
- Konuların hangi kelimelerle temsil edildiğini yorumlayın.