

# Dokümantasyon

- Haber kategorizasyonu 6 adet kategori(dünya, ekonomi, genel, güncel, kültür-sanat, magazin) ile yapılmıştır.
- Veri çekme işlemi hocaların ilettiği örnek siteler içerisindeki veriler kullanılarak gerçekleştirilmiştir.(13 kategoride 42 bin haber metni)
- Veriler çekilirken kategori sayısı 13'ten 6'ya düşürülmüştür.
- Dünya kategorisinden 1329; Ekonomi kategorisinden 1374; Genel kategorisinden 2398; Güncel kategorisinden 706; Kültür-Sanat kategorisinden 587; Magazin kategorisinden 505 adet veri çekilmiştir.

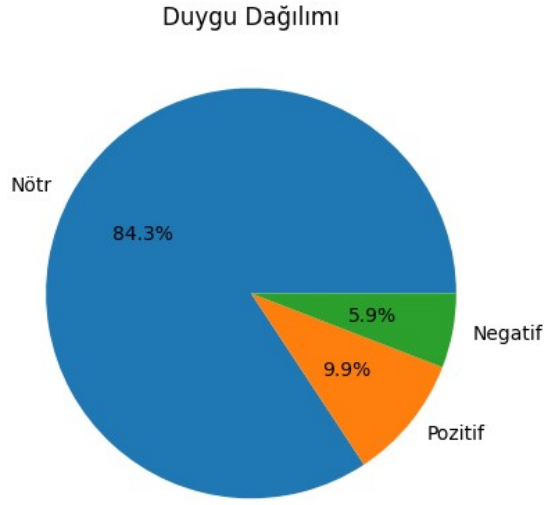
	Icerik	Kategori
0	'Ortak vizyonumuz var'\nDışışleri Bakanı Davut...	dunya
1	İsrail'den Gazze Şeridi'ne hava saldırısı\nİsr...	dunya
2	Cenaze için geniş güvenlik önlemleri alındı\nL...	dunya
3	Gözetimdeki sendikacılar serbest\nKKTC'de Sen...	dunya
4	Bisikletle Asya'da 3 bin kilometre yol katetti...	dunya
5	Ekonomik kriz, yabancı düşmanlığına yol açtı\n...	dunya
6	Putin, Amerikan karşıtı yasayı onayladı\nRusya...	dunya
7	'Suriye Ulusal Koalisyonu ile görüşmeye hazır...	dunya
8	İran'dan dev gözdağı\nBatılı ülkelerle nükleer...	dunya
9	İdam edecek ama organlara el koymayacak\nÇin'd...	dunya
10	Hız sınırı saatte 30 km olacak\nİspanya, şehir...	dunya
11	4 milyon dolar nafaka ödeyecek\nİtalya'nın esk...	dunya
12	Bağış'a : 'AB üyeliği kış ortasında yaz gecesi...	dunya
13	Mahmud Abbas'ın 'Bağımsız Filistin' umudu\nFil...	dunya
14	Yalçın'ın tahliyesi New York Times'da\nGazetec...	dunya
15	'Atina cami olmayan tek AB başkenti'\nBBC, Ati...	dunya

Şekil 1: İlk 16 Veri Seti

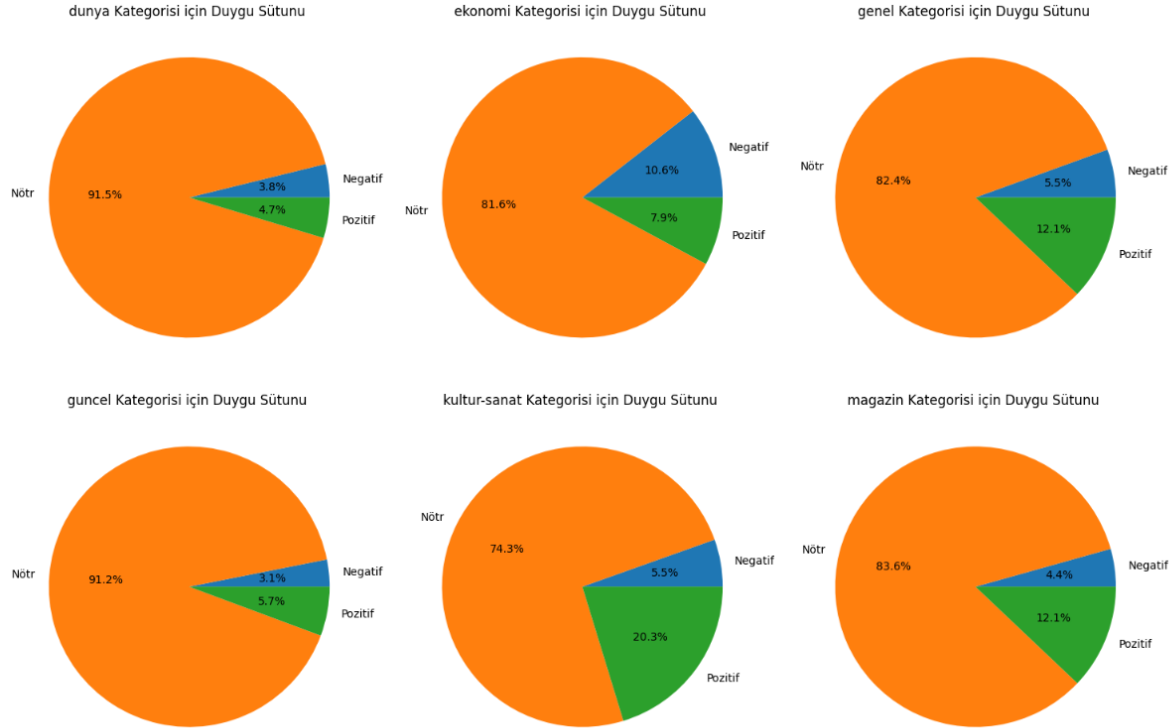
16	Brezilya'dan Filistin'e destek\nOrtadoğu'da ya...	dunya
17	Stolipin'in anıtına çelenk bıraktılar\nRusya D...	dunya
18	Hükümetin eğitim notu zayıf\nAlman Sendikalar ...	dunya
19	Kızıl Meydan'dan Türkiye'ye yılbaşı kutlaması...	dunya

Şekil 2: Kalan 4 Veri Seti

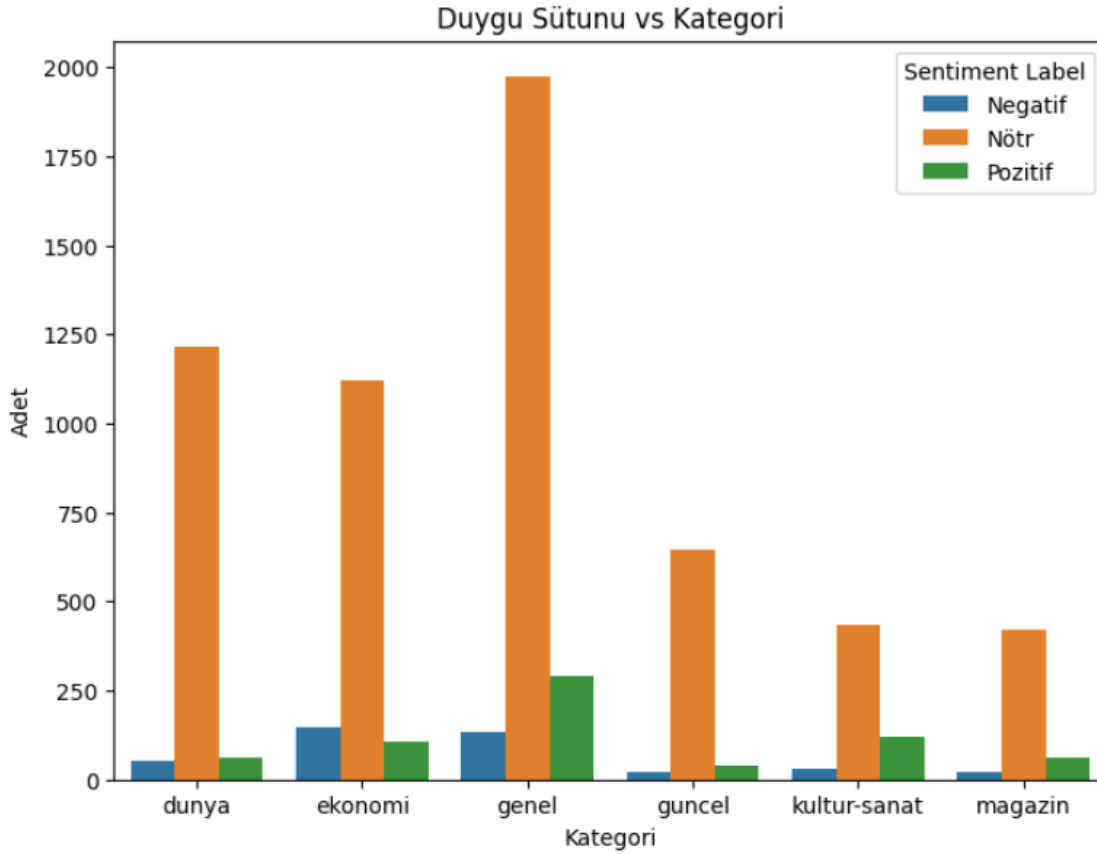
# Verilerin Görselleştirilmesi



Şekil 3: Tüm Verilerin Daire Grafiği



Şekil 4: Kategorilerin Daire Grafiği



**Şekil 5:** Kategorilerin Sütun Grafiği


- Haberin 6 kategorisi için de nötr ifadelerin yoğunlukta olduğu gözlemlendi.
- Genel kategorisinin diğer kategorilere nazaran en fazla pozitif ifade içeren kategori olduğu gözlemlendi.
- Ekonomi kategorisinin diğer kategorilere göre en fazla negatif ifade içeren kategori olduğu gözlemlendi.
- Güncel kategorisinin diğer kategorilere nazaran en az pozitif ifade içeren kategori olduğu gözlemlendi.
- Magazin ve güncel kategorisinin diğer kategorilere nazaran en az negatif ifade içerdiği gözlemlendi.
- Yüzdelik olarak kategoriler arasında nötr duygu durumu en fazla yoğunlukta olanın dünya kategorisi olduğu gözlemlendi.
- Yüzdelik olarak kategoriler arasında pozitif duygu durumu en fazla yoğunlukta olan kültür-sanat kategorisi olduğu gözlemlendi.
- Yüzdelik olarak kategoriler arasında negatif duygu durumu en fazla yoğunlukta olanın ekonomi kategorisi olduğu gözlemlendi.


# LSTM ve VADER Modelleri

## Arasındaki Farklar

- LSTM işlem yükü olarak VADER modelinden daha fazla işlemci yükü oluşturuyor.
- Tokenize ve mimari işlemleri LSTM de ekstra olarak yazılması gerekiyor. VADER'de tokenize işlemleri ve nöron oluşturma işlemleri modelin kendi içerisinde bulunuyor.

### Projenin colab linkleri:

 VADER.ipynb

 LSTM.ipynb

## Kaynakça

1. (<https://www.analyticsvidhya.com/blog/2022/07/sentiment-analysis-using-python/?ref=dailydev>)
2. (<https://www.kaggle.com/code/dimonyara/emotion-detection-from-tweets-roberta-fine-tuning>)
3. (<https://www.kaggle.com/code/dexterhari015/sentiment-analysis-nlp-emotion-detection>)