

# Metin Benzerliđi Hesaplama Projesi Raporu

## 1. Giriş

Bu proje kapsamında, haber metinleri kullanılarak metinler arası benzerlik hesaplanmıştır. Kullanılan veri seti, çeşitli haber başlıkları ve içeriklerinden oluşmaktadır. Amacımız, TF-IDF ve Word2Vec modelleri ile haber metinleri arasındaki benzerliđi sayısal olarak ölçmek ve farklı yöntemlerin başarımını karşılaştırmaktır.

## 2. Yöntem

Benzerlik hesaplamalarında iki temel yaklaşım uygulanmıştır:

1. TF-IDF (Term Frequency-Inverse Document Frequency): Haber metinleri vektörleştirilmiş ve kosinüs benzerliđi ile benzerlikler ölçülmüştür.
2. Word2Vec: Kelimeler vektörleştirilmiş, ardından her belge için ortalama vektör hesaplanarak kosinüs benzerliđi uygulanmıştır.

Word2Vec için farklı yapılandırmalarda CBOW ve Skip-gram mimarileri, pencere boyutu (3 ve 5) ve vektör boyutu (100 ve 200) gibi parametrelerle sekiz farklı model eğitilmiştir.

## 3. Sonuçlar ve Deđerlendirme

Her model için örnek bir haberin en benzer 5 haberi listelenmiş, benzerlik skorları tablo ve matrislerle gösterilmiştir.

- TF-IDF (Stemmed) ve TF-IDF (Lemmatized) modelleri özellikle başlık benzerliđi olan haberlerde başarılı sonuçlar vermiştir.
- Word2Vec modelleri, kelime vektörlerinden faydalananak anlam bazlı benzerliklerde öne çıkmıştır.
- Skip-gram mimarisi CBOW'a göre daha başarılı sonuçlar üretmiştir.
- Vektör boyutunun 200 olması, özellikle anlam benzerliđi açısından daha yüksek puanlar alınmasını sağlamıştır.

# Metin Benzerliđi Hesaplama Projesi Raporu

TF-IDF modelleri, daha yüzeysel ve hızlı hesaplamalar için uygunken, Word2Vec modelleri daha anlamsal benzerlikleri tespit edebilmektedir. Görsel olarak skorlar bar grafikleri ve benzerlik matrisleriyle sunulmuş, modeller arası tutarlılık Jaccard benzerliđi ile analiz edilmiştir.

## 4. Sonuç ve Öneriler

Genel olarak, TF-IDF yöntemi hızlı ve başarılı sonuçlar verirken; Word2Vec, anlamsal açıdan daha derin benzerlikler yakalayabilmektedir.

- TF-IDF, bilgiye dayalı kısa metin eşleşmeleri için uygundur.
- Word2Vec, haber gruplama, öneri sistemleri gibi uygulamalarda daha anlamlı eşleşmeler sağlar.

Gelecekte, BERT gibi bağlam-temelli dil modelleriyle daha güçlü sonuçlar elde edilebilir.