

Yerel Yonetim Butce Analizi

Ad Soyad: OMER KAYA

Ogrenci No: 2311081010

Bu projeyi hazirlarken yerel yonetimlerin faaliyet raporlarinda gecen harcama aciklamalarini anlam duzeyinde analiz etmeyi hedefledim. Amacim sadece kelime bazli degil, gercekten cumlelerin ne anlattigini kavrayabilecek yontemlerle benzerlik tespiti yapmakti. Veri setini olusturduktan sonra uyguladigim tum asamalari bu raporda detaylica paylastim.

1. Veri Seti

Projede temel olarak kullandigim veri seti, belediyelere ait faaliyet raporlarindan alinmis ornek cumlelerden olusmaktadir.

Her cumle bir harcama kalemini ifade etmekte ve toplamda 20 farkli ornek bulunmaktadir.

Bu cumleler uzerinden analizler gerceklestirilmistir.

2. On Isleme

Ham veriler uzerinde oncelikle kucuk harfe cevirme, noktalama isaretlerini kaldırma, durak kelimeleri temizleme gibi islemleri uyguladim.

Daha sonra ayni verinin iki farkli versiyonunu olusturdum: biri lemmatize edilmiş, diğeri ise stem edilmiş haliyle. Bu iki versiyon uzerinden paralel analizler gerceklestirdim.

3. Zipf Analizi

Zipf Yasasi'ni test etmek icin kelime frekanslarini log-log olcekte grafik haline getirdim.

Lemmatize ve stem edilmiş verilerde bu yasa buyuk olcude gecerli oldu. Grafiklerde frekanslarin duzgun bir sekilde azaldigi goruldu.

4. TF-IDF

TF-IDF yontemiyle her bir kelimenin belge icerisindeki onem derecesini hesapladim.

Bu yontem, nadir gecen ama anlam acisindan kritik olan kelimeleri on plana cikardi.

Her iki veri seti icin (lemmatize ve stem) ayri TF-IDF matrisleri olusturuldu.

Yerel Yönetim Bütçe Analizi

5. Word2Vec

Word2Vec modeli ile kelimeleri anlamsal olarak vektör uzayında temsil ettim.

Hem CBOW hem de Skip-Gram yöntemleriyle toplamda 16 farklı model oluşturdum.

Modeller, anlam olarak benzer kelimeleri ve cümleleri başarıyla tespit etti.

Özellikle CBOW daha stabil sonuçlar verirken Skip-Gram daha detaylı ilişkiler keşfetti.

6. Jaccard Benzerliği

Kelimeler üzerinden Jaccard benzerlik matrisi oluşturarak, cümleler arası yüzeysel benzerlikleri hesapladım.

Bu analiz, metinlerin sadece içerdiği ortak kelimelere dayandığı için anlamsal bağlantı kurmakta yetersiz kaldı.

Yine de genel uyumu görmek açısından faydalı oldu.

7. Genel Değerlendirme

Bu projede farklı vektörleştirme yöntemlerini karşılaştırmalı olarak kullanarak metin benzerliği analiz ettim.

TF-IDF yönteminin kelime bazında ise yaradığını, ancak anlami tam yakalayamadığını gördüm.

Word2Vec modeli ise daha isabetli sonuçlar verdi. Özellikle anlamsal yakınlıkları tespit etmekte çok başarılıydı.

Genel olarak elde ettiğim sonuçlar hem teknik açıdan tutarlı hem de anlamlıydı.