

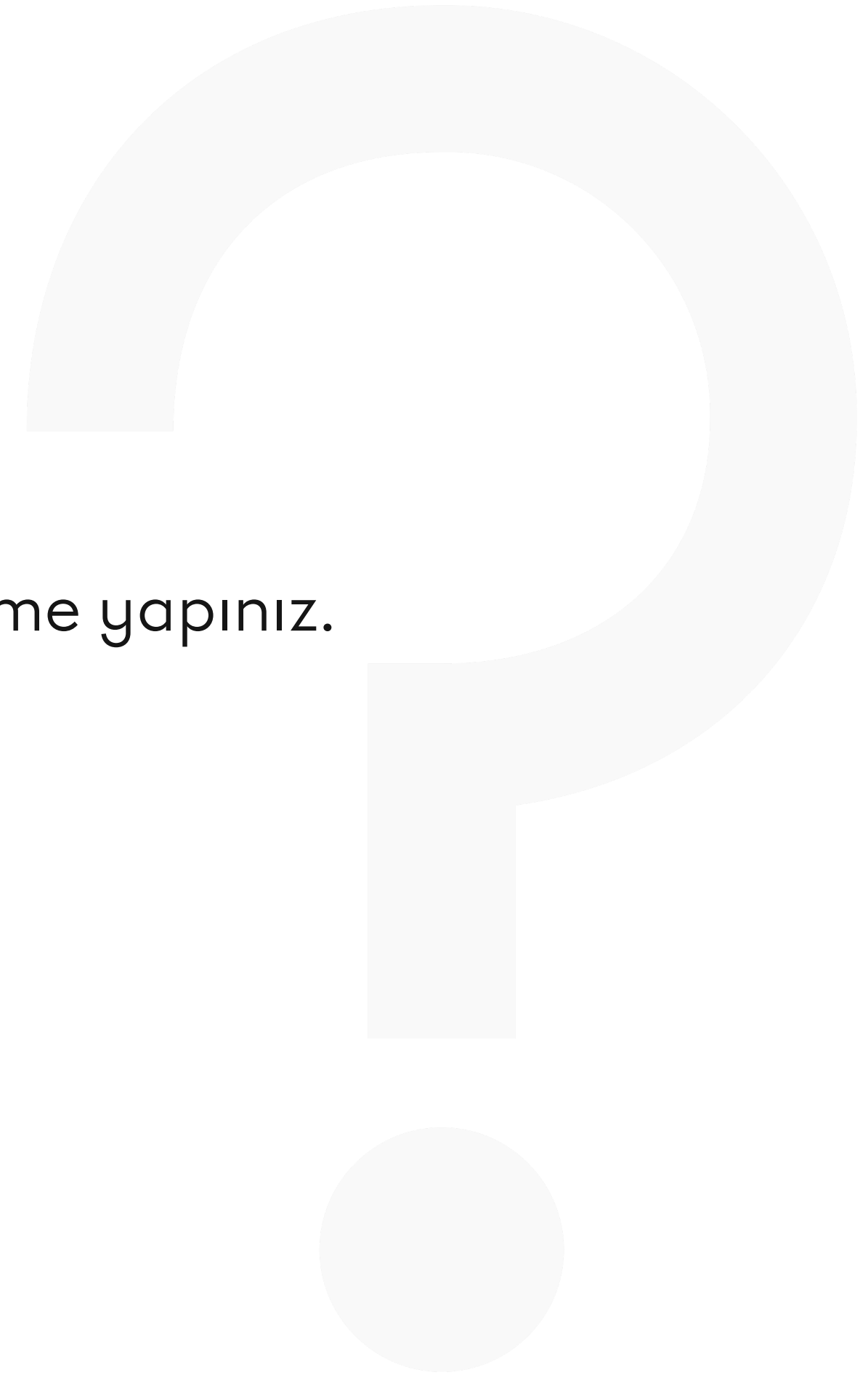


WIKIPEDIA

Metin Ön işleme ve Görselleştirme

Problem

Wikipedia metinleri içeren veri setine metin ön işleme ve görselleştirme yapınız.





Proje Görevleri

.....

Görev 1: Metin Ön İşleme İşlemlerini Gerçekleştiriniz

Adım 1: Metin ön işleme için clean_text adında fonksiyon oluşturunuz. Fonksiyon;

- Büyük küçük harf dönüşümü,
- Noktalama işaretlerini çıkarma,
- Numerik ifadeleri çıkarma İşlemlerini gerçekleştirmeli.

Adım 2: Yazdığınız fonksiyonu veri seti içerisindeki tüm metinlere uygulayınız.

Adım 3: Metin içinde öznitelik çıkarımı yaparken önemli olmayan kelimeleri çıkaracak remove_stopwords adında fonksiyon yazınız.

Adım 4: Yazdığınız fonksiyonu veri seti içerisindeki tüm metinlere uygulayınız.

Adım 5: Metinde az geçen (1000'den az, 2000'den az gibi) kelimeleri bulunuz. Ve bu kelimeleri metin içerisinden çıkartınız.

Adım 6: Metinleri tokenize edip sonuçları gözlemleyiniz.

Adım 7: Lemmatization işlemi yapınız.

Görev 2: Veriyi Görselleştiriniz

Adım 1: Metindeki terimlerin frekanslarını hesaplayınız.

Adım 2: Bir önceki adımda bulduğunuz terim frekanslarının Barplot grafiğini oluşturunuz.

Adım 3: Kelimeleri WordCloud ile görselleştiriniz.

Görev 3: Tüm Aşamaları Tek Bir Fonksiyon Olarak Yazınız

Adım 1: Metin ön işleme işlemlerini gerçekleştiriniz.

Adım 2: Görselleştirme işlemlerini fonksiyona argüman olarak ekleyiniz.

Adım 3: Fonksiyonu açıklayan 'docstring' yazınız.

miuul

miuul.com