

**BURSA TEKNİK ÜNİVERSİTESİ**  
**BİLGİSAYAR MÜHENDİSLİĞİ**

**PYTHON PROGRAMLAMA**  
**2024-2025 DÖNEM PROJESİ**

**HAZIRLAYANLAR:**

**RABİA KAMIŞ (-)**

**NAZMİ CİRİM(-)**

**HASAN GÜRFİDAN(-)**

**Proje Amacı ve Konusu:**

Bu projenin amacı, metin tabanlı veri kümeleri üzerinde Türkçe duygu analizi gerçekleştirmek için bir sistem geliştirmektir. Proje, kelimelerin pozitif, negatif veya nötr duygu değerlerini analiz ederek cümlelerin genel duygu durumunu belirlemeyi hedefler. Sistem, kelime tabanlı duygu sözlüğü, yoğunlaştırıcı ve zayıflatıcı kelimeler, olumsuzlama ifadeleri gibi temel verileri

kullanarak duygu analizi yapar. Bunun yanı sıra, metinleri kelimelere ayırma, stopwords filtreleme ve kök bulma işlemlerini Türkçe'ye özgü bir yaklaşımla gerçekleştirir. Analiz sonucunda, her bir cümle için duygu skoru hesaplanır, tahmin edilen sınıf belirlenir ve performans metrikleri (doğruluk, kesinlik, anma, F1-skoru) değerlendirilir.

### Kullanılan Teknolojiler:

1. **Python:** Projede Python, verilerin işlenmesi, metinlerin analiz edilmesi ve sonuçların hesaplanması için ana programlama dili olarak kullanılmaktadır.
2. **Pandas:** Excel dosyalarından veri okuma, veriyi düzenleme ve analiz sonuçlarını raporlama işlemlerinde kullanılmıştır.
3. **NLTK:** Metinleri kelimelere ayırmak (tokenization) ve Türkçe stopwords (anlamsız kelimeler) listesini almak için kullanılmıştır.
4. **Zemberek:** Metinlerin dilbilgisel analizini yapmak, kelimelerin köklerine indirgenmesi (stemming) ve anlamlı kelimeleri ayıklamak için kullanılmıştır.
5. **JPyype:** Python ile Java tabanlı kütüphaneler arasında köprü kuran bir kütüphanedir. Zemberek ile entegrasyon sağlamak için kullanılmıştır.
6. **XML ve ElementTree:** Lexicon (kelime listesi) dosyasını XML formatında okuyup işlemek için kullanılmıştır.

### Kodlar:

Kodlar tek bir Python dosyası yerine her işlevi ayrı bir dosya gerçekleştirecek şekilde birden fazla modüle ayrılarak yazılmıştır.

### İçeriği:

- Lexicon.py
- Preprocessing.py
- Sentiment\_analysis.py
- Metrics.py
- Main.py

### Lexicon.py

Lexicon ile ilgili kod parçaları, duygu analizi için temel verileri oluşturmak amacıyla tasarlanmıştır. `load_lexicon_from_xml`, kelimelerin duygu skorlarını içeren bir XML dosyasını okuyarak pozitif ve negatif skorların farkını hesaplar ve her kelime için bir genel duygu skoru oluşturur. Bu sözlük, metindeki kelimelerin duygu değerlerini anlamak için

bir temel sağlar. `load_modifiers` fonksiyonu, duygu yoğunluğunu artıran (ör., "çok") veya azaltan (ör., "biraz") kelimeleri bir metin dosyasından yükleyerek bir sözlük halinde döndürür. Bu bilgiler, kelime bazlı duygu etkilerinin analiz edilmesinde önemli bir rol oynar. `load_negation_words` ise, olumsuzluk kelimelerini bir liste olarak döndürür ve bu liste, metindeki kelimelerin duygu değerlerini ters çevirmek için kullanılır. Bu yapı, duygu analizinde olumsuzlukları, yoğunlaştırıcıları ve zayıflatıcıları dikkate alan esnek ve modüler bir temel oluşturur.

## Preprocessing.py

`tokenize_and_stem`, metni duygu analizi için hazır hale getiren kritik bir ön işleme fonksiyonudur. Bu fonksiyon, bir metni kelimelere ayırır (tokenization) ve Türkçe stopwords (önemsiz kelimeler) listesini kullanarak gereksiz kelimeleri filtreler. Daha sonra, Zemberek Türkçe doğal dil işleme kütüphanesini kullanarak kelimelerin köklerini (ör., "gidiyorum" → "git") tespit eder. Bu işlem, Türkçe gibi zengin bir dil bilgisine sahip bir dilde, metinlerin doğru şekilde analiz edilmesi için önemli bir adımdır. Ayrıca, alfanümerik olmayan karakterler ve noktalama işaretleri filtrelendir, böylece yalnızca analiz için anlamlı kelimeler işlenir. Bu yapı, metinlerin duygu analizi ve diğer doğal dil işleme görevleri için temiz, standartlaştırılmış bir veri seti sunar.

## Sentiment\_analysis.py

Dosya iki adet fonksiyondan oluşmaktadır; `calculate_sentiment` ve `analyze_sentiment`. Duygu analizi süreçlerini yöneten bu kod parçaları, metinlerin genel duygu durumlarını belirlemek için çalışır. `calculate_sentiment` fonksiyonu, bir metindeki kelimelerin duygu skorlarını toplar ve metnin genel bir duygu puanını hesaplar. Her kelimenin skoru, duygu sözlüğünden (lexicon) alınır ve güçlendirici (boosters) ya da zayıflatıcı (diminishers) kelimelerin etkileriyle çarpılarak güncellenir. Olumsuzluk kelimeleri (negation\_words) tespit edilirse, kelimenin skoru ters çevrilir (pozitifse negatif yapılır, negatifse pozitif yapılır). `analyze_sentiment` ise bu puanı alarak metni "Pozitif", "Negatif" veya "Nötr" olarak sınıflandırır. Bu yapı, kelime bazlı duygu analiziyle metnin genel duygusunu tahmin eden dinamik bir mekanizma sağlar.

## Metrics.py

Dosyanın içeriği 1 adet fonksiyondan oluşmaktadır, `calculate_metrics`, duygu analizi modelinin performansını ölçmek için geliştirilmiş bir fonksiyondur. Bu fonksiyon, analiz edilen verilerin gerçek sınıf değerleriyle tahmin edilen sınıf değerlerini karşılaştırarak doğruluk, kesinlik, anma ve F1-ölçütü gibi yaygın sınıflandırma metriklerini hesaplar. Öncelikle doğru pozitif (DP), yanlış pozitif (YP), yanlış negatif (YN) ve doğru negatif (DN) değerleri belirlenir. Ardından, bu değerler kullanılarak doğruluk oranı (modelin genel

başarısı), kesinlik (pozitif tahminlerin doğruluğu), anma (gerçek pozitiflerin doğru tahmin oranı) ve F1-ölçütü (kesinlik ve anma dengesini gösteren bir metrik) hesaplanır. Bu fonksiyon, büyük/küçük harf farklılıklarını düzeltir ve sıfır bölme hatalarına karşı koruma sağlar. Böylece, duygu analizi sisteminin hangi alanlarda başarılı veya zayıf olduğunu anlamak için kapsamlı bir performans değerlendirmesi sunar.

### Main.py Dosyası:

Bu Python dosyası, bir duygu analizi işlemi gerçekleştiren bir sistemdir. Dosyanın içeriği kullanıcının analiz etmek istediği cümleler ve bu cümlelerin doğru sınıflandırmalarından oluşur. Tüm işlemlerin yapıldığı dosyadır, diğer dosyalar ana dosya değil yardımcı ek işler yapan dosyalardır. Dosyada özel olarak kullanılan kelime dosyaları bulunmaktadır. Dosyalar sırasıyla işlevleriyle birlikte aşağıya yazılmıştır:

1. Lexicon.xml
2. Boosters.txt
3. Diminishers.txt
4. Negation\_words.txt
5. Ornek\_duygu-analiz-verisi-pozitif\_negatif.xlsx
6. Analyzed\_sentences.xlsx

1. Cümledeki kelimelerin pozitif negatif değerinin hesaplanması için lexicon.xml adlı dosyadan kelimelerin pozitif negatif ya da nötr değerleri çekilir ve en sonunda o değerler toplanarak mevcut cümlelerin puanı hesaplanır ve duruma göre değer negatifse cümle negatif, pozitif ise cümle pozitif cümle olduğuna karar verilir.
2. , 3. Duygu yoğunluğunu arttıran veya azaltan kelimelerin olduğu dosya.
4. Olumsuzlamaları tanımak için kullanılan kelimelerin olduğu dosya.
5. Dosyada yer alan her bir satır, analiz edilecek bir cümle ve bu cümlelerin doğru sınıflandırması hakkında bilgi içerir. Bu veriler, sistemin tahmin ettiği sınıflar ile karşılaştırılarak analiz performansı ölçülür.
6. Duygu analizi sonuçlarının kaydedileceği **çıktı dosyasını** ifade eder. Kod, analiz edilen her bir cümle için duygu puanını, tahmin edilen sınıfı ve diğer bilgileri bu Excel dosyasına yazar.

Bu kod, metin tabanlı veri kümeleri üzerinde duygu analizi gerçekleştirmek amacıyla tasarlanmıştır. İlk olarak, duygu analizi için gerekli olan kaynak dosyalar (duygu sözlüğü, yoğunlaştırıcı ve azaltıcı kelimeler, olumsuzlama kelimeleri) yüklenir. Ardından, giriş dosyası olarak belirtilen Excel dosyasındaki cümleler işlenir. Her bir cümle, önce kelimelerine ayrılarak (tokenization) ve köklerine indirgenerek (stemming) ön işleme tabi tutulur. Bu işleme, cümlelerin duygu sözlüğü ve yoğunluk kelimeleriyle karşılaştırılması izler; bu adımda pozitif ya da negatif kelimeler tanımlanır, olumsuzlama etkileri değerlendirilir ve duygu yoğunluğu artırılır ya da azaltılır. Tüm bu işlemler sonucunda

cümlelerin bir duygu puanı hesaplanır ve tahmini duygu sınıfı belirlenir. Analiz sonuçları, tahmin edilen sınıf, duygu puanı ve cümlelerin gerçek sınıfını içerecek şekilde bir Excel dosyasına kaydedilir. Kod ayrıca tahminlerin performansını değerlendirmek için doğruluk, kesinlik, duyarlılık ve F1-skoru gibi metrikler hesaplar ve bu metrikleri ekrana yazdırır. Bu sayede, analiz sonuçları detaylı bir şekilde incelenebilir ve sistem performansı ölçülebilir. Örneğin:

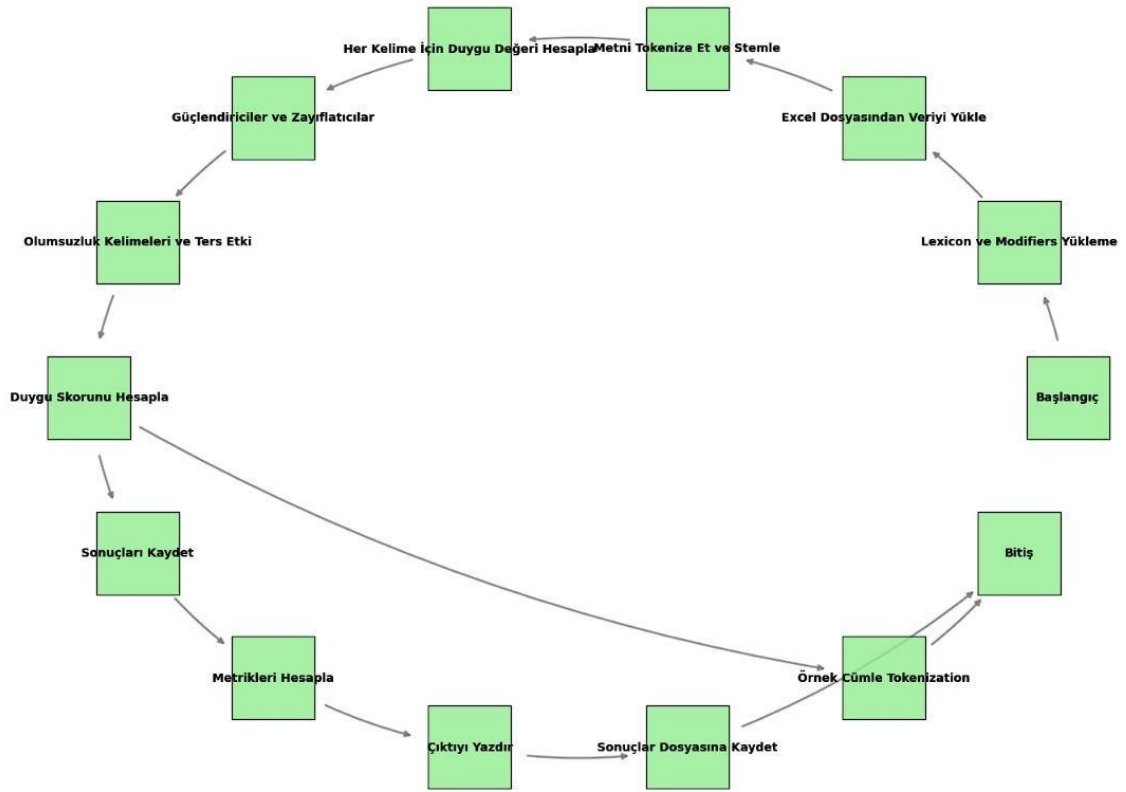
- Cümle: "Ben elma yemeyi seviyorum. "
  - Kelimelerine Ayrılmış Hali: "['ben', 'elma', 'yemeyi', 'sev']"
- ⑨ Gerçek Sınıf: Pozitif
  - ⑨ Tahmin: Pozitif
  - ⑨ Duygu Puanı: 0.6

#### Kod Çıktısı :

```
Doğruluk (Accuracy): 0.68
Kesinlik (Precision): 0.69
Anma (Recall): 0.76
F1-Ölçütü: 0.73
Analiz tamamlandı! Sonuçlar 'analyzed_sentences.xlsx' dosyasına kaydedildi.
Kelimelerine ayrılmış hali: ['ben', 'elma', 'yemeyi', 'sev']
```

#### Akış Diyagramı :

### Duygu Analizi Sistemi İş Akışı Diyagramı



### Sonuç:

Bu projede Python programlama dili kullanılarak metin tabanlı veri kümeleri üzerinde duygu analizi gerçekleştirilmiştir. Projede, duygu sözlüğü ve metin ön işleme teknikleri kullanarak cümlelerin duygu puanları hesaplanmış ve sınıflandırmalar yapılmıştır. Elde edilen sonuçlar, tahmin performansını ölçmek için doğruluk, kesinlik, duyarlılık ve F1-skoru gibi metriklerle değerlendirilmiştir. Projede elde ettiğimiz doğruluk oranı %68'dir. Bu oran, duygu analizi sistemi için temel bir başarı seviyesi sağlarken, geliştirilmeye açık alanların bulunduğunu da göstermektedir. Gelecekte, daha geniş veri kümeleri ve gelişmiş metin işleme teknikleri kullanılarak bu oranın artırılması hedeflenmektedir.