BURSA TEKNİK ÜNİVERSİTESİ

PYTHON DUYGU ANALİZİ ÖDEVİ

GRUP ÜYELERİ:

Fatih Parmaksız

Aytuğ Değer

Elif Pazarbaşı

Zehra Aktürk

Melike Badem

GitHub reposu:  
  
https://github.com/python-ekip/TurkishSentimentAnalysis

İçindekiler

[1. Projenin Kapsamı 3](#_Toc185892971)

[2. Veri kümesi 3](#_Toc185892973)

[3. Kod Açıklaması 3](#_Toc185892974)

[3.1 Kurallar 4](#_Toc185892975)

[4. Başarı Metrikleri 4](#_Toc185892976)

[5. Kaynakça 5](#_Toc185892977)

## Projenin Kapsamı

## Bu proje, Türkçe cümlelerin pozitif veya negatif duygu içerdiğini kural tabanlı bir yaklaşımla belirlemeyi amaçlamaktadır. Doğal dil işleme (NLP) alanında yaygın olarak kullanılan yapay zeka ve makine öğrenmesi tabanlı yöntemlerden farklı olarak, bu çalışmada dil bilgisi ve anlam analizine dayalı, manuel olarak tanımlanmış kurallar kullanılmıştır. Çalışmanın temel amacı, duygu analizi problemini şeffaf ve açıklanabilir bir yöntemle çözmektir.

Kural tabanlı yöntemler, belirli bir dildeki dil bilgisi yapılarının ve kelime türlerinin incelenerek anlam çıkarımı yapılmasını içerir. Türkçe’nin eklemeli bir dil olması, kelimelerin kök ve gövdelerinin analizini ve olumsuzluk eklerinin tespitini zorlaştırmaktadır. Bu bağlamda proje, kelime köklerini bulma, olumsuzluk eklerini tanımlama ve stopwords (anlam taşımayan bağlaçlar ve edatlar) gibi unsurları çıkarma adımlarını içermektedir. Ayrıca, belirli sıfat ve fiillerin polarizasyonu (pozitif ya da negatif anlam taşıması) temel alınarak cümlenin genel duygu durumu belirlenmiştir.

Proje sürecinde Zemberek, Argparse, Pandas kütüphanelerinden yararlanılmıştır.

## Veri kümesi

Proje kapsamında, ödev dosyasında sağlanan 804 cümlelik temel veri kümesi kullanılmıştır. Bu veri kümesi, kural tabanlı sınıflandırma sisteminin geliştirilmesi ve değerlendirilmesi için temel oluşturmuştur. Ek olarak, sistemin performansını daha kapsamlı bir şekilde test edebilmek için 200 negatif cümle ve 200 pozitif cümle manuel olarak oluşturulmuş ve test veri setine eklenmiştir. Bu sayede test veri seti, toplamda 400 cümleden oluşan dengeli bir yapı kazanmıştır.

Cümlelerin negatifliğini belirleyebilmek için:

93 adet negatif sıfat belirlenmiştir. Bu sıfatlar, cümlelerdeki olumsuz nitelikleri tespit etmek amacıyla kural tabanlı sistemde kullanılmıştır (*ör. üzgün, başarısız, kötü, zayıf*).

34 adet negatif fiil tanımlanmıştır. Bu fiiller, olumsuz eylemleri saptamak için kurallara dahil edilmiştir (*ör. kaybetmek, başaramamak, üzülmek*).

## Kod Açıklaması

Öncelikle noktalama işaretleri cümleden çıkarılır. Sonra cümleler kelimelere ayrılır. Zemberek kütüphanesi doğal dil analizi sırasında bu işlemleri gerçekleştirir. Kodda, TurkishMorphology ile yapılan dil analizi sürecinde, cümleler kelimelere ayrılır. Bu işlem şu kod ile gerçekleştirilir:

analysis = morphology.analyze\_sentence(sentence)

Zemberek, verilen cümleyi analiz ederek her kelime için ayrı bir analiz sonucu üretir. Bu, tokenization adımına karşılık gelir. Kod, Zemberek aracılığıyla kelimelerin türlerini (isim, fiil, sıfat vb.) belirler. Bu işlem, dil bilgisel analiz sırasında yapılır ve kelimenin gramer özelliklerini içerir.

after.best\_analysis()

Kodda, kelimelerin köklerini bulma işlemi Zemberek’in analiz aşamasında otomatik olarak yapılır. Her kelimenin kökü ve ek bilgisi analiz sonuçlarında yer alır. Cümledeki tüm fiiller (tumFiilleriBul) veya sıfatlar (sifatlariBul) bulunarak analiz edilir. Stopwords türünden kelimeler, anlam açısından genellikle etkisiz oldukları için bu analizlerde bir sonuca katkıda bulunmuyorlar. Daha sonra elde edilen yapı kurallar zincirinden geçirilir, negatif fiil, ne…ne yapısı, negatif sıfat, fiil sayısı, değişmem… yapısı gibi kuralar kontrol edilir. Eğer bir negatiflik varsa cümle negatif olarak işaretlenir.

taslak, çizim, diyagram, origami içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şekil 1.1

Şekil 1.1’de verilen kodun algoritmasını açıklayan bir akış şeması bulunmaktadır. Şemada, iki negatif fiilin arka arkaya gelmesi, negatif sıfat ve fiil kontrolü ile iki fiilin arka arkaya gelmesi durumlarında biri pozitif, diğeri negatif olduğunda gerçekleştirilen kontroller gösterilmektedir. Bu adımlar sonucunda sonuca ulaşılmaktadır.

Şekil 1.2

Kurallar tamamlanmadan önce sistemin doğruluk oranı %73 olarak tespit edilmiştir. Ancak, duygu analizine yönelik kuralların tamamlanması ve negatif fiil ile sıfatların analiz sürecine dahil edilmesi sonucunda bu oran Şekil 1.2 de görüldüğü gibi %83'e yükselmiştir. Söz konusu artış, sistemin Türkçe dil yapısına uyum sağlama kapasitesini artırdığı ve farklı duygu ifadelerini daha doğru bir şekilde analiz edebildiğini göstermektedir.

# 3.1 Kurallar

Bu kurallar, cümlede olumsuzluk belirten sıfatlar ve fiilleri tespit etmeye yönelik yazılmıştır. Özellikle, bir cümlede olumsuz sıfatlar (örneğin, "kötü", "negatif") veya fiillerin olup olmadığını kontrol eder. Ayrıca, fiillerin köklerinin olumsuzluk içerip içermediği de incelenir. Bu kurallar, dildeki olumsuzluk anlamlarını belirlemek için kullanılır. Cümlede ironi olup olmadığını kontrol eden bir kural bulunur. Bu, cümlenin anlamını çelişen bir şekilde ifade edip etmediğini anlamaya yöneliktir. Örneğin, cümlede "(!)" işaretinin bulunması, ironi kullanımını işaret eder. Aynı şekilde, cümledeki fiil ya da sıfatın olumsuz olup olmadığı da kontrol edilir. Bu kurallar, cümledeki olumsuzlukları ve ironiyi anlamaya yönelik kritik işlevler görür.

Bir cümledeki fiillerin olumsuz olup olmadığını belirlemek için bir dizi kontrol yapılır. Bu kurallar, fiilin olumsuzluk ekleri taşıyıp taşımadığını inceler ve dildeki negatif anlamların doğru bir şekilde yorumlanmasını sağlar. Değişmem…. Ve ne…ne için de kural yazılmıştır. Bu kuralın amacı şu tip cümlelerin duygu tespitini sağlamak içindir:

“Saçının bir telini değişmem Ankara'ya.”

“Vize alamadıkları için düğününe ne kardeşi ne kuzeni katılabildi.”

Bu kurallar, belirli kelimelerin dildeki anlamını analiz eder. Örneğin, "değişmem" kelimesinin pozitif anlamda mı yoksa negatif anlamda mı kullanıldığını kontrol eder. Aynı şekilde, "Ne ... ne ..." yapısının kullanımı da analiz edilerek, dildeki olumsuzluk veya anlam kayması tespit edilmeye çalışılır.

Tüm bu kuralların birleşimiyle, bir cümlenin dilbilgisel yapısı daha doğru bir şekilde analiz edilir. Cümlenin içindeki fiillerin sıfatlarla olan ilişkisi, negatif yapıların varlığı ve genel anlam çözümlemesi bu kurallar aracılığıyla sağlanır. Bu, dilin anlamını ve yapısını doğru bir şekilde yorumlamaya yardımcı olur.

## Başarı Metrikleri

Excel dosyasındaki cümleler ve bunların doğru sınıflandırmaları (gerçek sınıflar) pandas kütüphanesi ile okunur. df adlı veri çerçevesi, her bir satırda bir cümle ve ona karşılık gelen doğru sınıf etiketini içerir. Her cümle için, classify\_sentence\_with\_reason fonksiyonu çalıştırılır. Bu fonksiyon, verilen cümleyi dilbilgisel analiz yaparak, pozitif veya negatif olarak sınıflandırır ve sınıflandırma sebebini belirler. Bu, modelin tahmini sınıfıdır.

Her cümle için gerçek sınıf etiketi (true\_class) ile tahmin edilen sınıf (predicted\_class) karşılaştırılır. Bu karşılaştırma, eğer tahmin doğruysa (true\_class == predicted\_class), doğru tahmin sayısını artırır. Dosyadaki tüm cümleler üzerinden yapılan tahminlerin doğruluk oranı şu şekilde hesaplanır :

metin, yazı tipi, ekran görüntüsü, beyaz içeren bir resim

Açıklama otomatik olarak oluşturuldu

Doğru tahminlerin sayısı, toplam cümle sayısı ve doğruluk oranı, her bir cümlenin sınıflandırılmasıyla birlikte bir log dosyasına kaydedilir. Ayrıca, doğruluk oranı ve diğer istatistikler, bir metin dosyasına yazılır. Bu dosya, modelin performansını gösteren rapor olarak kaydedilir.

## Kaynakça

1. <https://github.com/loodos/zemberek-python>
2. Chen, D. Y. (2017). *Pandas for everyone: Python data analysis*. Addison-Wesley Professional.
3. Rasche, H. (2023). Python-Argparse.