

Türkiye Açık Kaynak Platformu  
Online Yarışma Programı

# Türkçe Doğal Dil İşleme

- TAKIM ADI -

mAln

Takım Lideri: Kevser Cansu YURDUSEVER

## - EKİBİMİZ -

- Ekip 1 üyeden oluşmaktadır. Takım kaptanı Kevser Cansu YURDUSEVER'dir
- 1998 yılında Trabzon'da doğdum. KTO Karatay Üniversitesi Bilgisayar Mühendisliği bölümünden bu yıl(2022) mezun oluyorum. Üniversitede 3. sınıfta Yapay Zeka ve Veri Bilimi Çalışma grubuna katıldım ve halen aktif olarak çalışmalarına devam etmekteyim. Bu grupta yapay zeka ve veri bilimi alanında kendimi geliştirdim.
- Bu süreçte 'TÜBİTAK 1005' programına grup arkadaşlarımla katıldım. Aynı projede önerdiğimiz yöntem için patent başvurusunda bulunduk. Ayrıca bu grubun üyelerinden oluşan bir takımla TeknoFest'e Akıllı Ulaşım kategorisinde katıldık. Projemizle finale kaldık. Yarışma süreci devam etmektedir.
- **Bu takımdaki görevim proje geliştirme, model oluşturma ve kodlamadır.**

## - EKİP ÜYELERİNİN PROJEYE SUNDUĞU KATKI -

- Ekipteki tek üye olarak araştırma, proje geliştirme ve kodlama görevlerim vardı.
- Projenin planlanması, üzerinde çalışılacak verinin oluşturulması ve projede bahsi geçen deyimler ve atasözlerinin duygu skorlarını içeren bir veri setinin hazırlanması aşamalarında görev aldım.

## - PROBLEM -

- Projenin odağı Türkçe Doğal Dil İşleme'de atasözleri ve deyimlerimiz için duygu skorları içeren bir veri seti bulunmamasıdır. Bunun altında yatan temel sebep ise bağlamın anlaşılamamasıdır. Deyimler ve atasözleri mecaz ve ironi içerdiğinden içeriğin anlaşılması zorlaşmaktadır.
- Türkçe dilimizde de duygu analizi yapılmak istendiğinde alınan sonucun en doğru şekilde olması adına atasözleri ve deyimler için bir veri seti ihtiyacı vardır.

## - PROBLEMİN ÇÖZÜMÜ -

Probleme çözüm önerisi olarak herhangi bir Türkçe metinde gerçekleştirilecek duygu analizinde kullanılmak üzere **atasözleri ve deyimlerin duygu skorlarını içeren** bir veri seti oluşturulması planlanmıştır.

# -YÖNTEM-

- Projede kullanılan yöntemde deyim ve atasözleri yerine anlamlarına duygu analizi yapılması planlanmıştır.
- Türkçe deyim ve atasözlerinin anlamları çevrimiçi ortamdan elde edilerek bir veri seti oluşturulacaktır.
- ‘sentistrength’ kütüphanesi kullanılarak atasözleri ve deyimlerin anlamlarına cümle bazlı analiz yapılarak duygu skorları elde edilecektir.
- Ardından Türkiye Türkçesindeki Deyimlerinde Duyguların Analizi(AVCI, Yıldız Yenen, 2015) adlı makaleden yararlanılarak kendi oluşturduğum veri setinde de bulunan deyimlerin etiketlenmesi yapılacaktır. Negatif ve Pozitif şeklinde etiketleme yapılacaktır.
- Analiz sonucunda elde edilen skorlar ile etiketler karşılaştırılarak doğruluk oranı incelenecek ve iyileştirmeler yapılacaktır.

## -TEKNİK ÇALIŞMALAR-

- Çevrim içi olarak elde edilen veri seti .xlsx ve .csv formatında kaydedilmiştir.
- Duygu analizi için çeşitli kütüphaneler araştırılmıştır. Bunlar SentiTurkNet, KeNet gibi synsetler; sentistrength, scikit-learn, zeyrek, Turkish Stemmer, nltk ve spaCy'dir.
- Bunlarla çeşitli vectorization, stemming ve lemmatization işlemleri denenmiştir.
- Öncelikle atasözü ve deyimlerin Türkçe anlamları lemma'lara ayrılarak SentiTurkNet synset'inde karşılık gelen lemmaların skorları elde edilmiştir. Ancak bu synset'te tekrar eden lemmalara rastlanmıştır. SentiTurkNet synset'inde düzenlemeler yapılması gerekmektedir. Buna ek olarak cümleleri kelime bazlı değerlendirirken anlamın tamamen karşılanmadığı da fark edilmiştir.
- Bu nokta da cümle bazlı analiz yapılmasına karar verilmiştir.



## -TEKNİK ÇALIŞMALAR-

- Cümle bazlı analizde kullanılmak üzere yabancı bir kaynak kullanılmıştır: 'sentistrength' kütüphanesi.
- Bu kütüphanenin kullanılması için Türkçe deyim ve atasözleri İngilizce'ye çevrilmiştir. Ardından kütüphanede ilgili metotlar kullanılarak bunların duygu skorları çıkarılmıştır.
- Son olarak veri setinde negatif 'N', pozitif 'P' şeklinde etiketleme yapılmıştır.
- Etiketlemenin ardından duygu analizi sonucu çıkan skorlar ve etiketlerin karşılaştırılması yapılmıştır.

## - PROJE İŞ AKIŞI -

# -YOL HARİTASI-

Problemden çözüme giden yol, çözümden gelecekte planlanan hedefler

# -DEMO VIDEO LİNKİ-

## - KAPANIŞ -