

TÜRKÇE DOĞAL DİL İŞLEME

# Türkçe Doğal Dil İşleme

8 - 9 Ağustos 2024

ProfPlay



# FATİH ÇÖREKÇİ

Bu yıl tek kişilik kadro olarak burdayım.

## Eğitim Hayatı

- 2024 - İstanbul Üniversitesi AUZEF Yönetim Bilişim Sistemleri
- 2005 - Marmara Üniversitesi Psikolojik Danışmanlık ve Rehberlik

## Çalışma Hayatı

- 2005-2015 - MEB Atama
  - (Yüksekova-Taşköprü-Salihli)
  - (ilköğretim, Meslek Lisesi, RAM, BİLSEM)
- 2015 - devam Salihli Bilim ve Sanat Merkezi (PDR - Müdür - PDR)

## ProfPlay TDDİ Model Çalışması

Türkçe Doğal Dil İşlemeye katkı sağlamak amacı ile proje çalışmasına başladım. Duygu analizi ile farklı alanlardaki müşteri (faydalanan/hedef kitle) yorumlarını üretime (hizmete) dahil etmeyi hedefliyorum.



## ProfPlay TDDİ Model Çalışmasının Sağladığı Çözümler

Yorum alan şirket/şahıs için ürün/hizmetinin memnuniyet düzeyini tespit etmesi, yorumlarda rakipleri ile kıyaslanılmada kendi yerini görmesini sağlayacak başlangıç aşaması bir çalışmadır.

# PROJE İŞ AKIŞI

Yarışma startı verildikten sonra Çalışma süresini 4 (dört) ana bölüme ayırdım.

Her bir bölüm 5 x 3 = 15 birimlik parçalara ayırdım.

	1 Ana Bölüm					2 Ana Bölüm					3 Ana Bölüm					4 Ana Bölüm				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Yarışma Şartnamesi ve Dokümanlarını inceleme	■					■					■					■				
Literatür taraması		■	■				■					■								
Akademisyen görüşlerinin alınması		■	■									■								
Örnek çalışmaların incelenmesi			■	■			■	■				■	■			■	■			
İlgili videoların, seminer ve webinarların izlenmesi		■		■			■			■	■			■	■		■			
Yöntem ve kütüphane, çalışma alanı tespiti				■	■	■														
web scraping, kaggle Dataset oluşturma/bulma, temizleme, ön işleme					■	■			■	■										
Datasetten model oluşturma, github													■	■		■			■	
Trail, test çalışmaları ve doğruluk oranını yükseltme																■	■			
Kontrol ve tekrar planlama							■				■				■		■	■		■
İstenilen veri ile çalışma ve Sunum																			■	■



## &lt;VERİ SETİ&gt;

Kullanılan veri seti için öncelikle sosyal medya üzerinden web scraping çalışmaları yapıldı. Kaggle platformundan ilgili veri setleri tarandı. Son olarak mağaza yorumlarından duygu analizi veri setinde (11429) karar kılındı. 3 null veri temizlendikten sonra Olumlu (4252), Olumsuz(4237) ve nötr(2937) olarak etiketlenen veri seti ön işlemeden geçirildi. Veri seti train(0.8), test(0.2) olarak ayrıldı. Bert tr kullanılarak modellenen veri setinden cümle vektörleri çıkarıldı.

## &lt;YÖNTEM VE TEKNİKLER&gt;

Google Colab üzerinde çalışmalar yürütüldü.

Başlangıçta web scraping ve API ile twitterdan veri çekme çalışmaları yürütülmüştür. Bu çalışma ile istenilen düzeyde sonuç elde edilememiştir. Kaggle platformu üzerinden alınan veri seti ön işleme sonucu 11426 yorum ile çalışıldı. csv dosyası google drive'a yüklendi.

stefan-it Turkish Bert modelinden faydalanıldı.

Son olarak jupiter nootbook github hesabına yüklenmiştir.

İlgili kaynaklar:

<https://github.com/akoksal/BERT-Sentiment-Analysis-Turkish/blob/master/PyIstanbul%20Notebooks/BERT%20Features.ipynb>

<https://github.com/stefan-it/turkish-bert>



## &lt;MODEL EĞİTİMİ VE DEĞERLENDİRME&gt;

Model eğitilirken bert-base-turkish-128k-uncased kullanıldı.  
%80 train ve %20 test olmak üzere dataframe random(42) ikiye ayrıldı.

Çıkan modelin sonuçları

Olumlu:      Olumsuz:      Nötr:

Ortalaması :



## &lt;SONUÇLAR&gt;

Kaggle platformundan elde edilen alışveriş duygu analiz (olumlu/olumsuz/nötr) dataseti ile eğitilmiş bir model ortaya çıktı. Bu model ile yorumun muhatabı olan firma/şahıs tespiti ve doğruluğu %80 üstü duygu analizi hedeflenmiştir.

## &lt;PROJE YOL HARİTASI&gt;

Özellikle sosyal medya yorumları ile yeni datasetler oluşturmak.  
Duygu analizi doğruluğu yüksek, doğru muhatabı tespit edebilen  
modeller oluşturmak. Sahte veya bot yorumları eleyebilmek.  
Eğitim alanında kalitenin artırılması için paydaş duygu analizi  
sonuçlarının değerlendirilmesi ile paydaş memnuniyeti anketlerine  
yeni bir soluk getirmek.



## <DEMO VİDEO>

Projenin demo videosunun ve linkinin eklenmesi.

# TEŞEKKÜRLER