

Türkiye Açık Kaynak Platformu  
Online Yarışma Programı

# Türkçe Doğal Dil İşleme

# – TAKIM ADI –

Takım Adı: VV Cephei A  
Proje Adı: Covid-19 Haberleri

# — EKİBİMİZ —

## Ayşe Yağmur AYYILDIZ

Erciyes Üniversitesi Bilgisayar Mühendisliği 3. Sınıf öğrencisiyim. Web geliştirme üzerine çalışmalarda bulunmaktayım. Görüntü işleme ve yapay zekâ üzerine TÜBİTAK projeleri geliştirmekteyim. TÜBİTAK 2209-A ve 2242 kapsamında 2 projem destek almaktadır. TEKNOFEST 2020 yarışmalarında 2 kategoriden finale kalmış bulunmaktayım. Ekip içerisinde yazılım geliştirici olarak görev aldım.

## İlknur İMAMHALİLOĞLU

Erciyes Üniversitesi Bilgisayar Mühendisliği 3. Sınıf öğrencisiyim. Web ve mobil geliştirme üzerine çalışmalarda bulunmaktayım. Görüntü işleme ve yapay zekâ üzerine TÜBİTAK projeleri geliştirmekteyim. TÜBİTAK 2209-A ve 2242 kapsamında 2 projem destek almaktadır. TEKNOFEST 2020 yarışmalarında 2 kategoriden finale kalmış bulunmaktayım. Ekip içerisinde yazılım geliştirici olarak görev aldım.

## Hatice AKSOY

Erciyes Üniversitesi Bilgisayar Mühendisliği 3. Sınıf öğrencisiyim. Yapay zekâ alanında doğal dil işleme, görüntü işleme, derin öğrenme ve makine öğrenmesi konuları ile ilgilenmekteyim. Aynı zamanda web ve masaüstü uygulamaları ile de ilgilenmekteyim. TÜBİTAK 2209-A kapsamında desteklenen projem bulunmakta. TEKNOFEST 2020 için 2 kategoriden finale kalmış bulunmaktayım. Ekip içerisinde yazılım geliştirici olarak görev aldım.

## Buket BEKER

Erciyes Üniversitesi Bilgisayar Mühendisliği 1. Sınıf öğrencisiyim. Dil işleme, görüntü işleme ve web geliştirme konularına ilgili ve bu konular üzerinde projeler geliştirmekteyim. TEKNOFEST 2020 yarışmalarına tarım kategorisinden başvurmuş bulunmaktayım. Ekip içerisinde yazılım geliştirici olarak görev aldım.

# — PROBLEM NEDİR? —

Dünya, son zamanlarda hayatımızı büyük ölçüde etkileyen bir virüs ile karşı karşıya kalmıştır. Uzun bir süredir hem dünyada hem ülkemizde devam eden Covid-19 pandemisi hakkında birçok haber sitelerinde ve sosyal medya hesaplarında bilgi paylaşımları yapılmaktadır. Bu paylaşımlarda bilgiler bazen karmaşık ve anlaşılmaz olmakla birlikte bazen haberin özüne inmeden önce genel anlamda birbirini tekrar eden uzun bilgiler içermektedir. Bu bilgiler doğrultusunda ise bireylerin vakit kaybı yaşamasına ve birçok kaynaktan farklı türde birçok bilgi edinilmesinden dolayı bilgi kirliliğine neden olmaktadır. Bu nedenle güncel bilgilerin uzunluğu okuyucuların ilgisini ve okuma isteğini olumsuz yönde etkilemekte ve yanlış bilgiler edinmesine sebep olmaktadır.

## – ÇÖZÜM NEDİR? –

Dünya'nın ve ülkemizin son zamanlarda maruz kaldığı Covid-19 virüsüne yönelik sosyal medya ve haber siteleri başta olmak üzere birçok platformda bilgi paylaşımları yapılmaktadır. Bu bilgilerin çokluğu, başta bilgi kirliliği olmak üzere bilgilerin doğruluğunun bilinmemesi, bütün bilgileri okumaya çalışırken oluşan zaman kaybı gibi sorunları ortaya koymaktadır. Bu tür oluşabilecek zaman kaybı ve bilgi kirliliğinin önüne geçilmesi amacıyla bir web sitesi tasarlanmıştır. Yapılan bu web sitesinde, başta Sağlık bakanlığının resmi web sitesinden ve ülkemizde görev alan başlıca 3 büyük haber sitesinden veri kazıma yöntemiyle çekilen haberlerin arka planda oluşturulan algoritma sayesinde özetleri çıkarılarak doğru bilgiye en kısa sürede ulaşılabilmesi sağlanmıştır. Bunun yanında sadece Türkiye gündeminin dışında dünyada olan gelişmelerin de takip edilebilmesi için ayrı bir sayfada dünyadaki haberlerin özetlenmesi gerçekleştirilmiştir. İlaveten haber sitelerinin pandemi sürecinde sosyal medyadan aktif bir şekilde yaptığı paylaşımları da göz önünde bulundurarak her haber sitesinin resmi Twitter hesaplarından bilgileri çekerek web sayfasında gösterilmesi sağlanmıştır. Bu sayede birçok farklı platformdaki bilgilere tek bir yerden en azami sürede ulaşılmasına olanak verilmiştir.



## – İŞ AKIŞI –

Bu sistemin geliştirilmesi için Python dili ve nltk, Math, Numpy, Flask, Request, beautifulsoup4 kütüphaneleri kullanılmıştır. İlk olarak tüm sistemin kullanıcılara sunulacağı web sitesinin tasarımı HTML&CSS kullanılarak tasarlanmıştır. Sitenin tasarımı bittikten sonra Python programlama dili kullanılarak, veri kazıma yöntemi ile verilerin kazınacağı sitelerden haberler çekilmiştir. Çekilen haberler tarafımızca geliştirilen özet çıkarma algoritması ile yazılmış olan kod yardımıyla özetlenmiştir. Özeti çıkarılan haberler ayrı section alanlarında gösterilmiştir. Kullanıcının isterse haberin tamamını görmesi için her haber özetinin sonuna haber linki konularak asıl haber sayfasına ulaşılması sağlanmıştır. Tasarlanan web sitesi 5 farklı sayfadan oluşmaktadır. Ana sayfa içeriğinde Sağlık Bakanlığı sitesinden çekilen Türkiye'deki günlük vaka sayıları, geçmiş günlerdeki günlük ve aylık vefat sayıları vb. güncel verilerin bulunmasının yanında, Sağlık Bakanı Fahrettin Koca'nın resmi Twitter hesabından paylaştığı tweetler de bulunmaktadır. Tasarlanmış olan diğer 4 sayfada ise Sabah, Milliyet, Hürriyet gibi yerli gazeteler ve BBC News gazetesindeki haberlerin özetlendiği ve kullanıcıya sunulduğu bir sayfa bulunmaktadır. Her sayfada haberlerin özetlenmesinin yanında her haber sitesine ait resmi twitter hesaplarından da anlık olarak paylaşılan tweetlerin çekilip sayfaya entegre edilmesi sağlanmıştır. Tüm sistem bileşenlerinin birleştirilmesi ile proje tamamlanmıştır.

# – DEMO GÖSTERİMİ – İLE İLGİLİ VIDEO LINKİ

[https://drive.google.com/file/d/1trPaavRziMhilGp3dx\\_eeFXSGyHI6QVy/view?usp=sharing](https://drive.google.com/file/d/1trPaavRziMhilGp3dx_eeFXSGyHI6QVy/view?usp=sharing)

# — KAPANIŞ —

Yarışmayı düzenleyen Türkiye Açık Kaynak Platformu Türkçe Doğal Dil İşleme Çevrim İçi Yarışma'sını düzenleyen ekibe, jüri ve mentör üyelerine ve proje geliştirme aşamasında bize desteklerini esirgemeyen mentörümüz Kemal Can Kara'ya çok teşekkür ederiz.

Dinlediğiniz için teşekkürler.