



Keşf-i Lisan

Türkçe Doğal Dil İşleme ile Covid-19 Gündem Analizi

Kürşat Cakal

- Gazi Üniversitesi- Bilgisayar Mühendisliği



- Baykar Savunma – Ar Ge Mühendisi
 - Robotik
 - Gömülü Sistemler
 - Yapay Zeka



- TÜBİTAK - Teknofest'19 Akıllı Şehirler ve Ulaşım
Türkiye Birinciliği



- Microsoft - Açık Akademi Robolife



- Gazi Üniversitesi Otonom Araç Geliştirme Ekibi
 - Teknofest'18 Robotaksi – ileri (Finalist)
 - Nvidia Racecar Türkiye Yarışması – (Finalist)



Can Korkut

- Gazi Üniversitesi Bilgisayar Mühendisliği



- Sestek – Ar Ge Mühendisi
 - Yapay Sinir Ağları
 - Makine Öğrenimi
 - Derin Öğrenme



- TÜBİTAK - Teknofest'19 Akıllı Şehirler ve Ulaşım
Türkiye Birinciliği



- Cumhurbaşkanlığı Fikir Maratonu İkincilik.



- Gazi Üniversitesi Otonom Araç Geliştirme Ekibi
 - Teknofest'18 Robotaksi – ileri (Finalist)
 - Nvidia Racecar Türkiye Yarışması – (Finalist)



Problem Nedir?

Değişen Gündem

Covid-19 Pandemisi ile alakalı küresel ve ulusal gündem hızla değişmektedir.

Artan Haber Sayısı

Artan kaynak ve haber sayısı gündem analizini zorlaştırmaktadır.

Gündem Takip Zorluğu

Covid-19 pandemisi harici farklı temalarda da gündem takibi gerçekleştirmek ciddi zaman gerektirmektedir.

Yoğun Tempolu Hayat

Günlük hayattaki yoğunluk gündem takibini zorlaştırmaktadır.

- Gündemi gün boyunca bizim yerimize takip eden bir bot olsa nasıl olurdu?
- Geliştirdiğimiz, Keşf-i Lisan Covid-19 BOT 7/24 gündem takibi gerçekleştirerek hedef anahtar kelimelerin yer aldığı metinleri toplayıp, sınıflandırıp, analiz ederek gündem takibi gerçekleştirir.
- Bu sayede nitelikli bilgiye erişim cazip bir hale gelmektedir.

Akış Şeması

1 1 0 1 0 1 0 0 0 1 0 1 0 0 1 0 1 1 0 0
1 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 1 0 0
1 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 1 0 0
1 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 1 0 0
1 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 1 0 0
1 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 1 0 0
1 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 1 0 0



Veri Toplama



Yazar
(Yakında)

Özet

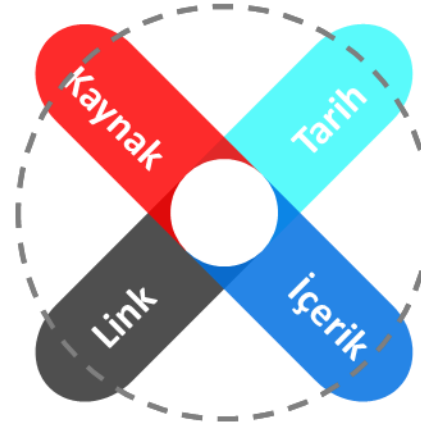


Kümeleme



Resimler
(Yakında)

Abstractive
(Yakında)



Normalizasyon



Analiz



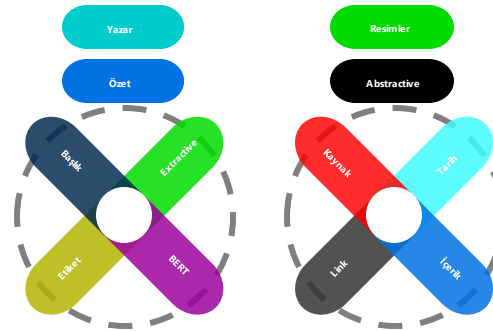
- BERT
- Abstractive
- Extractive



COVID-19
Veri Seti

Akış Şeması

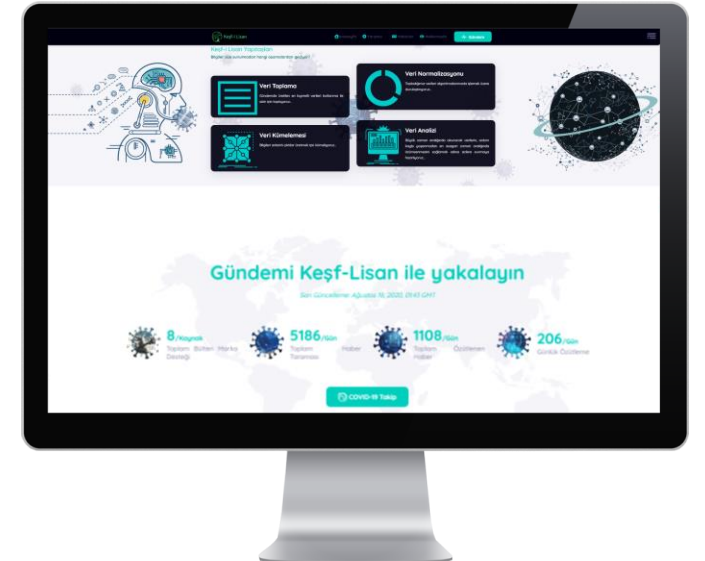
Covid -19 Veri Kümesi



Veri Tabanı



Uygulama





Scrap.py



Kullanılan Teknolojiler

- Link toplama için;
 - Google API
 - Httplib2
- Veri Kazıma
 - BeautifulSoup

Koronavirüs aşısında umutlandıran gelişme!

Tüm dünya koronavirüs salgınının pençesinde boğuşurken herkes Covid-19'dan kurtulmamızı sağlayacak aşı çalışmalarının bitmesini bekliyor. Son gelişmelere göre önce ABD'den sonra da Almanya'dan aşı konusunda umut veren haberler geldi.

Tüm dünya nefesini tutmuş [Covid-19](#)'u yenebilecek bir aşının ortaya çıkmasını bekliyor. Dünya genelinde koronavirüs vaka sayısı 19 milyon 200 bini, virüs yüzünden hayatını kaybedenlerin sayısı ise 717 bini geçmiş durumda.

Karayipler'deki Barbados ulkesi, tatile gelen yabancı turistlerin 1 yıl boyunca uzaktan çalışarak adada kalması için esine rastlanmayan bir program 01.01.2021 itibari ile baslatıyor."



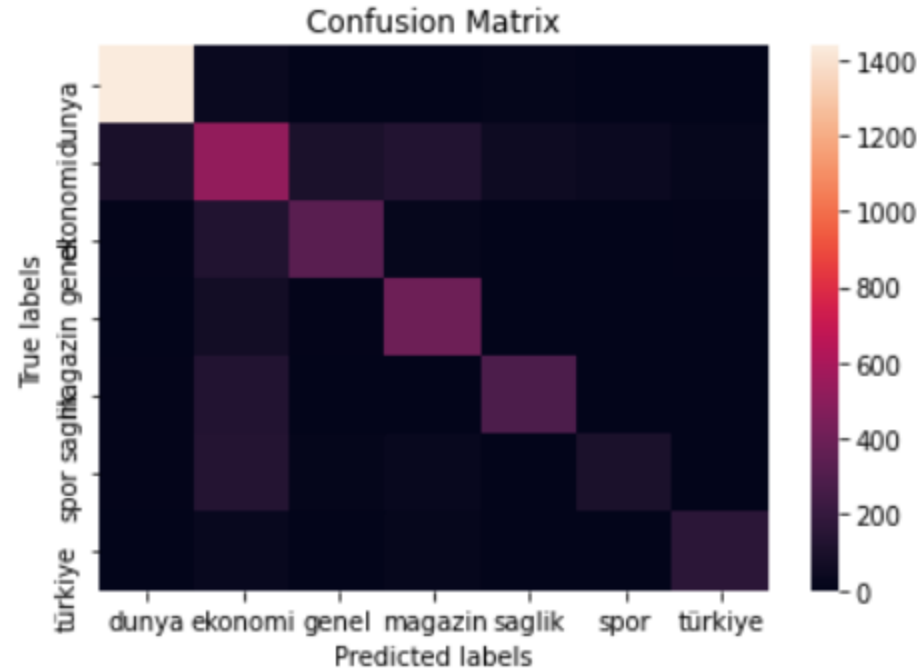
Normalizer.py



karayiplerdeki barbados ulkesi tatile gelen yabancı turistlerin bir yıl boyunca uzaktan çalışarak adada kalması için esine rastlanmayan bir program bir ocak ikibinyirmibir itibari ile baslatıyor .

- Noktalama işareti ayıklama, - Büyük küçük harf normalizasyonu, - Tarih, - Saat, - Rakam

- Haber sınıflandırılma gerçekleştirilmesi için etiketli verilerden oluşan veri kümesi kullanarak geliştirilen Derin öğrenme modeli ile sınıflandırma modeli oluşturulmuştur.
- Veri Seti: Türkçe 42 bin Haber veri kümesi (Açık Kaynak)
- Kullanılan Araçlar: PyTorch, TorchText



Türkiye, koronavirüs salgını ile mücadele eden ülkelere yardım eli uzatmaya devam ediyor. Şimdiye kadar bir çok ülkeye tıbbi malzeme yardımında bulunan Türkiye, salgın ile mücadele eden Kazakistan Cumhuriyeti'ne tıbbi malzeme ve ekipman yardımında bulundu. Cumhurbaşkanı Recep Tayyip Erdoğan'ın talimatıyla Sağlık Bakanlığı'nca hazırlanan yardım malzemeleri, Ankara Esenboga Havalimanı'nda getirilerek uçağa yüklendi. Yardım malzemeleri arasında 30 ventilatör cihazı, 50 bin N95 maske, 100 bin cerrahi maske, 35 bin tulum, 2 bin siperlik, 2 PCR cihazı, 30 bin SARS-CoV-2 (2019-nCov) RT-qPCR Tespit Kiti, 30 bin Viral Nükleik Asit İzolasyon Kiti ve 10 bin kutu Hidroksiklorokin tablet yer aldı. Havalimanındaki törene Dışişleri Bakan Yardımcısı Yavuz Selim Kıran ve Kazakistan'ın Ankara Büyükelçisi Abzal Saparbekuly katıldı. 'KAPIMIZI ÇALAN 152 ÜLKENİN 140'INA TIBBİ YARDIMDA BULUNDUK'Bakan Yardımcısı Kıran, mart ayından bu yana dünyanın görünmeyen düşmanla mücadele ettiğini belirterek, Covid-19 ile mücadelede alınan tedbirler yereldir. Yerel anlamda her ülke kendine özgü tedbirleri almaktadır. Ama bu mücadelede başarı gösterilmesinin yolu, küresel iş birliği ve dayanışmadır. Bugün kapımızı çalan 152 ülkenin 140'ına tıbbi yardımda bulunan bir ülke olduğumuzun altını çizmek istiyorum. İhtiyaç halinde Kazakistan ve Kazakistan halkının yanında olmaya devam edeceğiz. Bu kara günde Türkiye Cumhuriyeti halkı 7 den 70'e unsurlarıyla birlikte Kazakistan halkının yanında olmuş, bu kara günde dayanışmasını bir kez daha göstermiştir diye konuştu.Yapılan konuşmaların ardından yardım uçağı Kazakistan'a havalandı.

Şimdiye kadar bir çok ülkeye tıbbi malzeme yardımında bulunan Türkiye, salgın ile mücadele eden Kazakistan Cumhuriyeti'ne tıbbi malzeme ve ekipman yardımında bulundu. Bu kara günde Türkiye Cumhuriyeti halkı 7 den 70'e unsurlarıyla birlikte Kazakistan halkının yanında olmuş, bu kara günde dayanışmasını bir kez daha göstermiştir diye konuştu.Yapılan konuşmaların ardından yardım uçağı Kazakistan'a havalandı.

Nasıl Çalışır ?

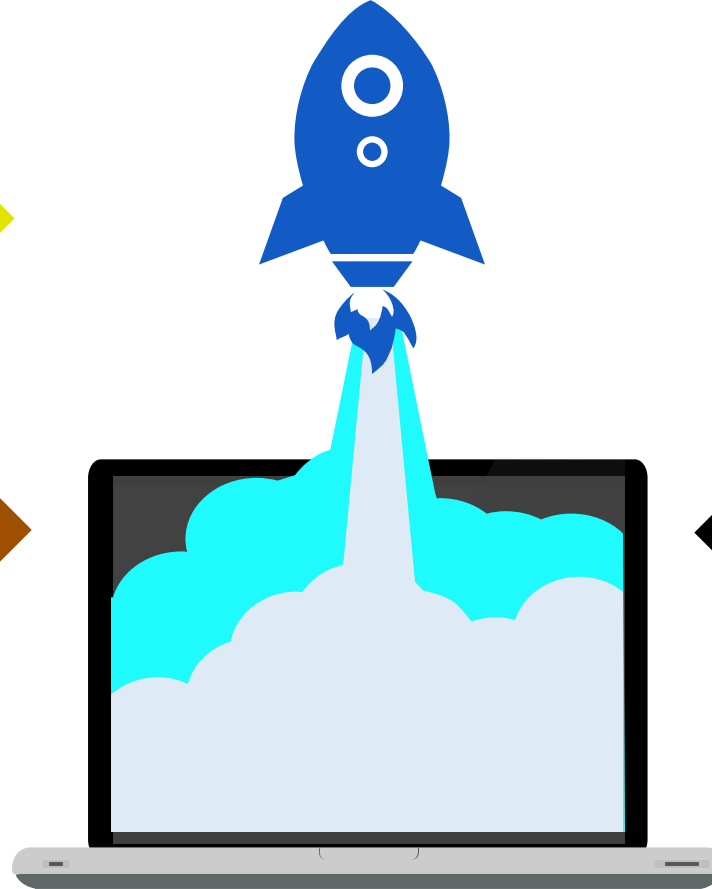
- İçerik cümlelerini ayrıştırır.
- Cümle içeriklerindeki sıklık kontrolü ile matris yapısı oluşturur.
- Oluşturulan matris yapısını derecelerine göre işler.
- Cümlelere ait derece değerlerini içeren matrisi oluşturur.
- Derece matrisi ve belirlenen offset değerine göre istenilen sıklıkta özet elde edilir.

Scarp.py
Haber sitelerinden metin toplama aracı

Normalizer.py
Metin Normalizasyonu

2

1

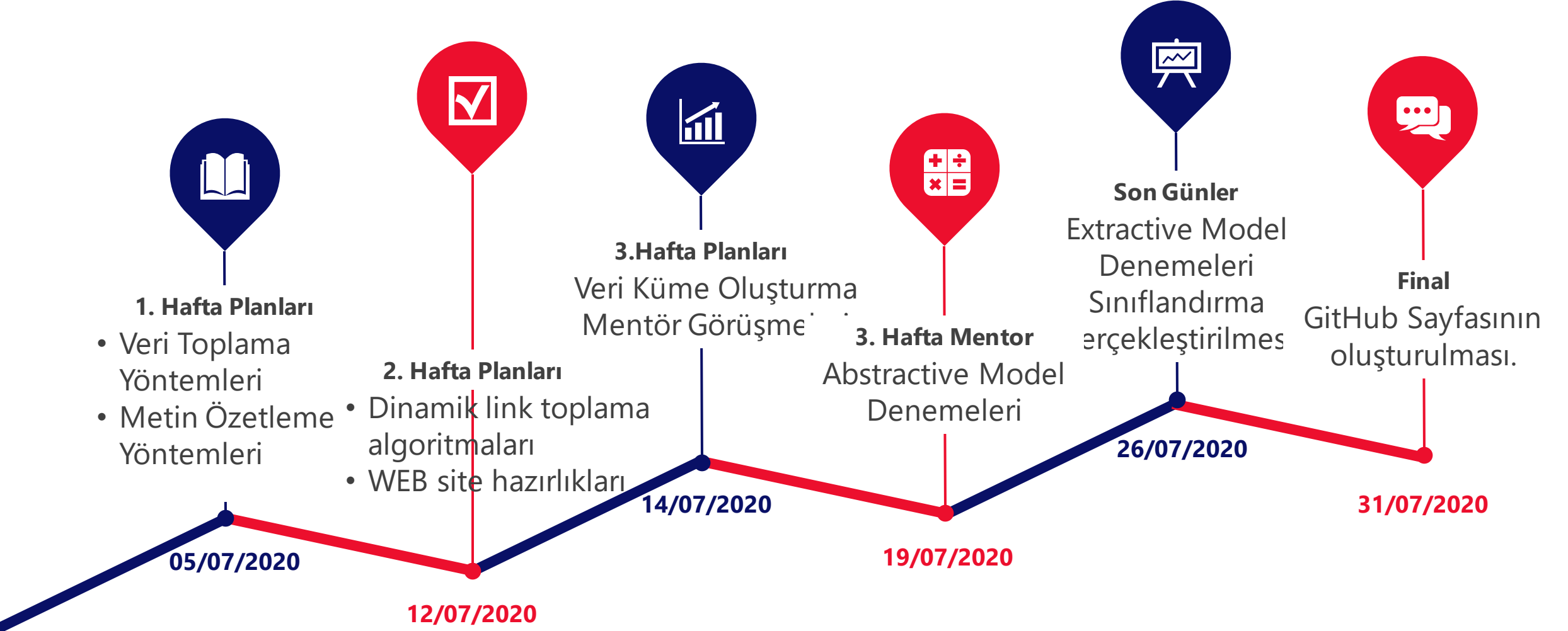


3

4

Classifier.ipynb
PyTorch ile metin sınıflandırma demosu

Extractive_Sum.py
Extractive metin özetleme yöntemi



İş Paketi	Gerçekleştiren(ler)
Veri Toplama	Can Korkut – Kürşat Cakal
Normalizasyon	Can Korkut
Extractive Özetleme	Kürşat Cakal
Sınıflandırma – PyTorch	Can Korkut
WEB Sitesi	Kürşat Cakal
Abstractive Araştırmaları	Can Korkut – Kürşat Cakal

Veri Toplama

Toplamda yaklaşık 1.200 adet korona virus haber metni toplandı.

Sınıflandırma

42bin haber veri kümesi kullanılarak %72 başarı oranında 7 kategoride haber sınıflandırılması gerçekleştirildi.

Özetleme

İki farklı yaklaşım ile toplanan metinler özetlendi.

WEB Sitesi

Kullanıcı arayüzünü oluşturan korona virus web sitesi demosu oluşturuldu.

Uygulama Demosu



<https://www.youtube.com/watch?v=OZ5Cs1swN3A>

<http://kesf-i-lisan.co/>



Keşf-i Lisan



Teşekkürler...