

TÜRKÇE DOĞAL DİL İŞLEME SENARYOSU

TÜRKÇE DOĞAL DİL İŞLEME

8 - 9 Ağustos 2024


KARAYEL

Ekibimiz



Serkan Dereli

Takım Danışmanı

Doç. Dr. Öğretim
Üyesi



Murat

Takım Kaptanı

Mekatronik
Mühendisliği
4. Sınıf



Berra Nur

Takım Üyesi

Bilgisayar
Mühendisliği
3. Sınıf



Caner Furkan Karali

Takım Üyesi

Bilgisayar
Mühendisliği
3. Sınıf

Hedef Tabanlı Duygu Analizi

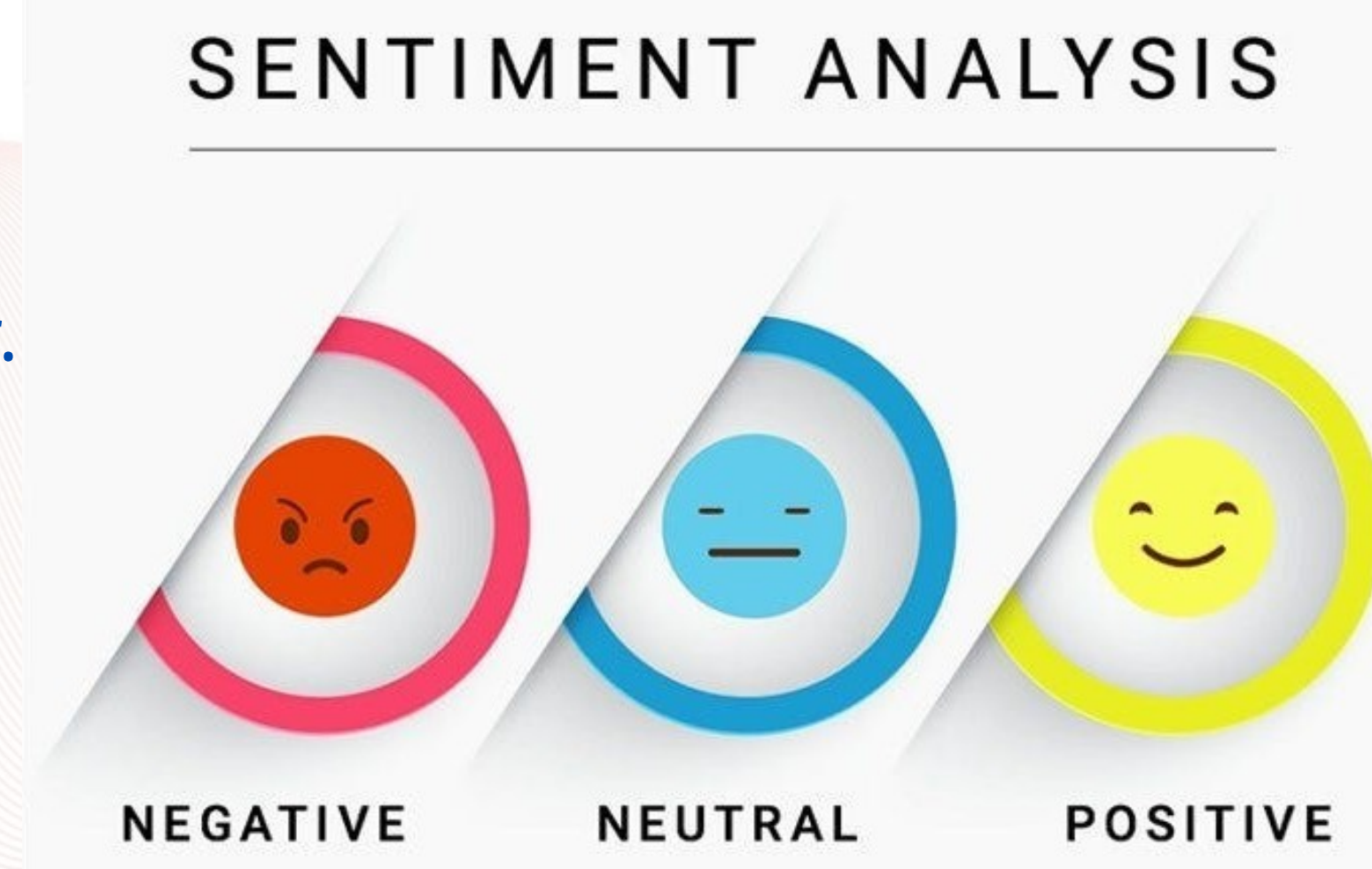
aspect-based sentiment
analysis

Duygu analizi yaygın olarak kullanılan bir DDİ görevidir.

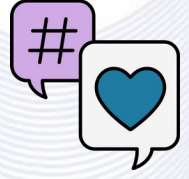
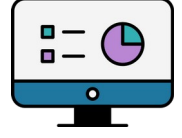
Hedef tabanlı duygu analizi ise belirli bir konu veya hedefe yönelik duygu durumunu

belirler
Metin içerisindeki belirli öznelerle ilgili
duygu

ifadelerini ayırt eder ve sınıflandırır.
İşletmelerin stratejik kararlar almasına
yardımcı olan veri odaklı bir tekniktir.



PROJEMİZİN SAĞLADIĞI KATKILAR NELERDİR?



Eğitilen modeller ve geliştirilen algoritma ile uçtan uca şikayet duygu analizi



Şikayetlere hızlı ve etkili geri dönüt

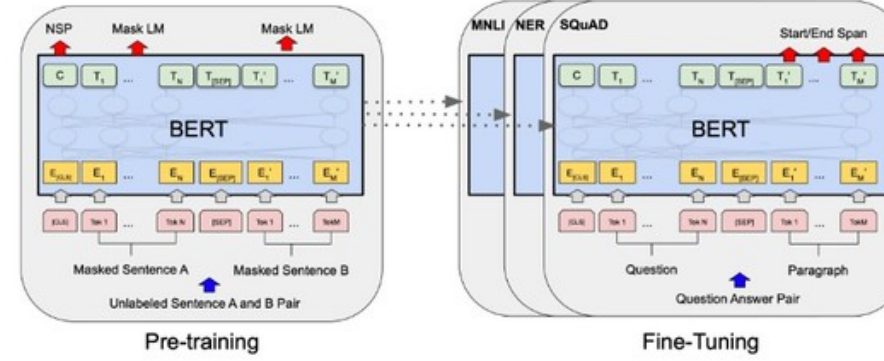


Geliştirilen araçlar ve yayınlanan veri setleri ile akademik ilerleme sağlanması



Şikayet analizi gereksinimlerinin karşılanması

<PROJE İŞ AKIŞI>



Veri Kazıma ve Etiketleme
İşlemleri

Test Ortamının
Hazırlanması

Model Eğitimi



WEB SİTE



WEB SCRAPING



DATA

Prediction

POST /predict/ post_predict_

Metin içinde organizasyon (B-ORG) etiketli varlıklar bulun ve her birinin duyguyu analizi yapılır.

Parameters

body * required Example Value Model

object (body)

{ "text": "Türkiye SuperOnline hizmetinden memnundur"

Parameter content type application/json

Veri Kazıma ve Etiketleme Süreçleri

Etiketleme Alanı

1 haftadır internet sorunu yaşıyoruz pazartesi günü (29.07.2024) arıza kaydı oluşturdum, 24 saat içinde çözülecek demelerine rağmen hiç ilgilenilmedi. Bugün internet tamamen kesildi defalarca arıza kaydı oluşturmamıza rağmen gelen giden olmadı! Turkcell pişmanlıktır. Yıllardır fiber internet diyerek yüksek faturalar alan **Turkcell** daha yeni fiber hizmeti vermeye başladı bu konu içinde ayrıca şikayetim olacak!

Yeni Etiket Ekle

Pozitif

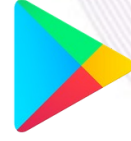
Nötr

Negatif

Kaydet

Tümünü Kaydet ve İndir

şikayetvar



ekşisözlük

Karayel-Uren

Tweet Scraping Visualization Edge Extension

Şikayetvar

Marka (örn: turkcell-bulut)

Başlangıç Sayfası (örn: 1)

Bitiş Sayfası (örn: 350)

Verileri çek Durdur Kaydet Yükle

Metin

x.com şunu diyor:

Scraped Tweets:

sadece bir el valorant oynamak istedim ilk superonline gitti simdi elektrikler bana hzur yokkk

#superonline #turkcell internet yok. Bizler 2 gün fatura ödemesek diye kesersiniz siz sebepsiz yere interneti kestiniz nasıl olacak bu ??

Turkcell Superonline internetimde son 1 haftadır geceleri sürekli bağlantı kopmaları oluyor ve bugün bağlantı kopmasıyla da kalm. 1000Mbps hız paketi olan internetimi birden 60mbps vermeye başladı ne müşteri hizmetlerine bağlanabiliyorum edebiliyorum. @Turkcell

KARAYEL

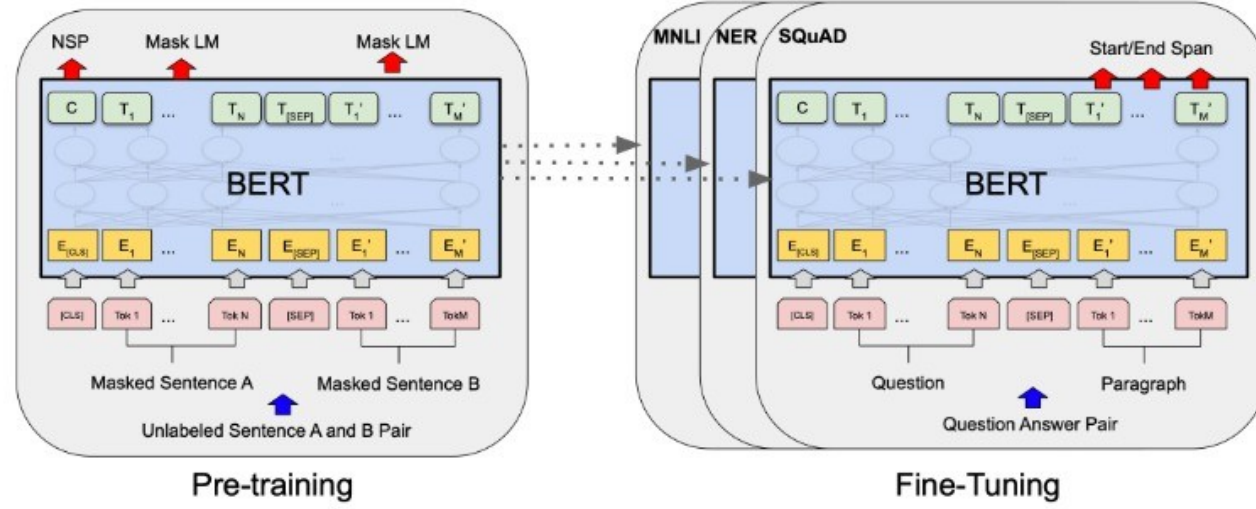
Tweet
Scraper

Açık sekmenizdeki
yüklenmiş tweetleri çeker

Tweetleri Çek

Farklı kaynaklardan veri kazıma süreçlerini hızlandırmak adına bir araç geliştirdik.

Varlık-İsim Tanıma (NER)



Veri Kazıma ve Etiketleme Süreçleri

Tokens	Bilişim	Vadisi	yerleşkesindek i	Teknofest	DD İ	Yarışması	oldukça	iyi	ilerliyor
Tags	1	2	0	1	2	2	0	0	0
Polarities	2	2	-1	2	2	2	-1	-1	-1

Oluşturduğumuz veri setimiz ile, BERTurk modeli kullanarak bir NER modeli eğittik.

Bu modelimiz ile birlikte müşteri şikayetinde bulunan markaları tespit ettik.

Tespit ettiğimiz varlık isimleri hakkında yukarıda görüldüğü şekilde bir encoding işlemi yaptık.

Sonrasında ABSA model eğiterek analizi gerçekleştirdik.

MODEL EĞİTİMİ VE DEĞERLENDİRME

```
learning_rate = 2e-5
per_device_train_batch_size = 16
per_device_eval_batch_size = 16
num_train_epochs = 3
weight_decay = 0.01
metric_for_best_model = "eval_accuracy"
```

Transformers kütüphanesi, Trainer nesnesi kullanarak 'dbmdz/bert-base-turkish-uncased' modelini sol tarafta belirtilen parametrelerle eğittik.

Train ve eval metrikleri aşağıdaki gibi görülebilir.

```
***** train metrics *****
epoch                =          3.0
total_flos           = 1714660GF
train_loss           =          0.0104
train_runtime        = 0:02:24.54
train_samples_per_second =          61.414
train_steps_per_second  =           3.84
```

```
***** eval metrics *****
epoch                =          3.0
eval_accuracy        =          0.9806
eval_f1 score        =          0.7113
eval_loss            =          0.0856
eval_precision       =          0.6641
eval_recall          =          0.7658
eval_runtime         = 0:00:01.67
eval_samples_per_second =          221.489
eval_steps_per_second  =          14.367
```


SONUÇLARIMIZ

Düşük parametrelili GPT-2 modeli denenmiş ve alınan sonuçların yetersizliği üzerine BERTturk ile NER ve ABSA olmak üzere iki farklı model eğitilmiştir.

ORGANISATION

LOCATION

DATE

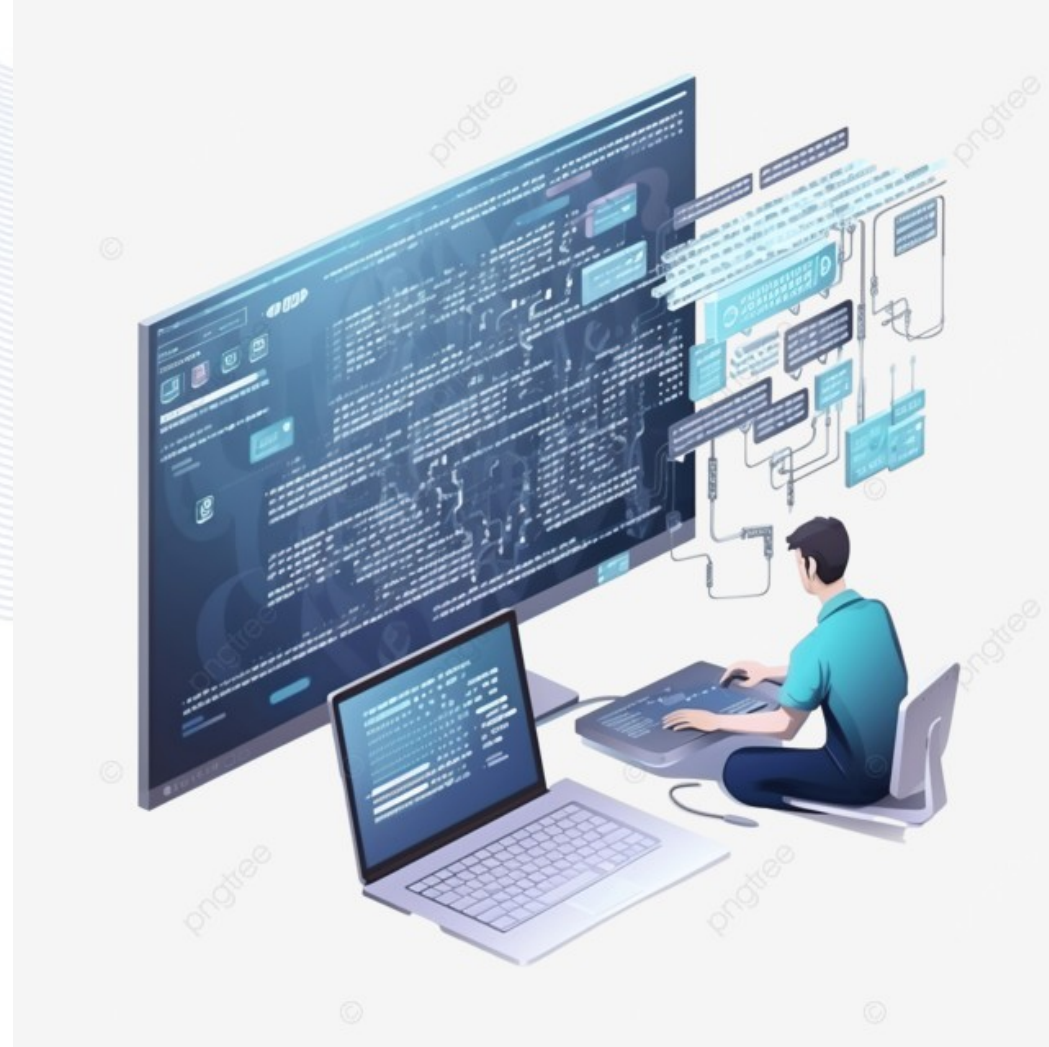
PERSON

WEAPON

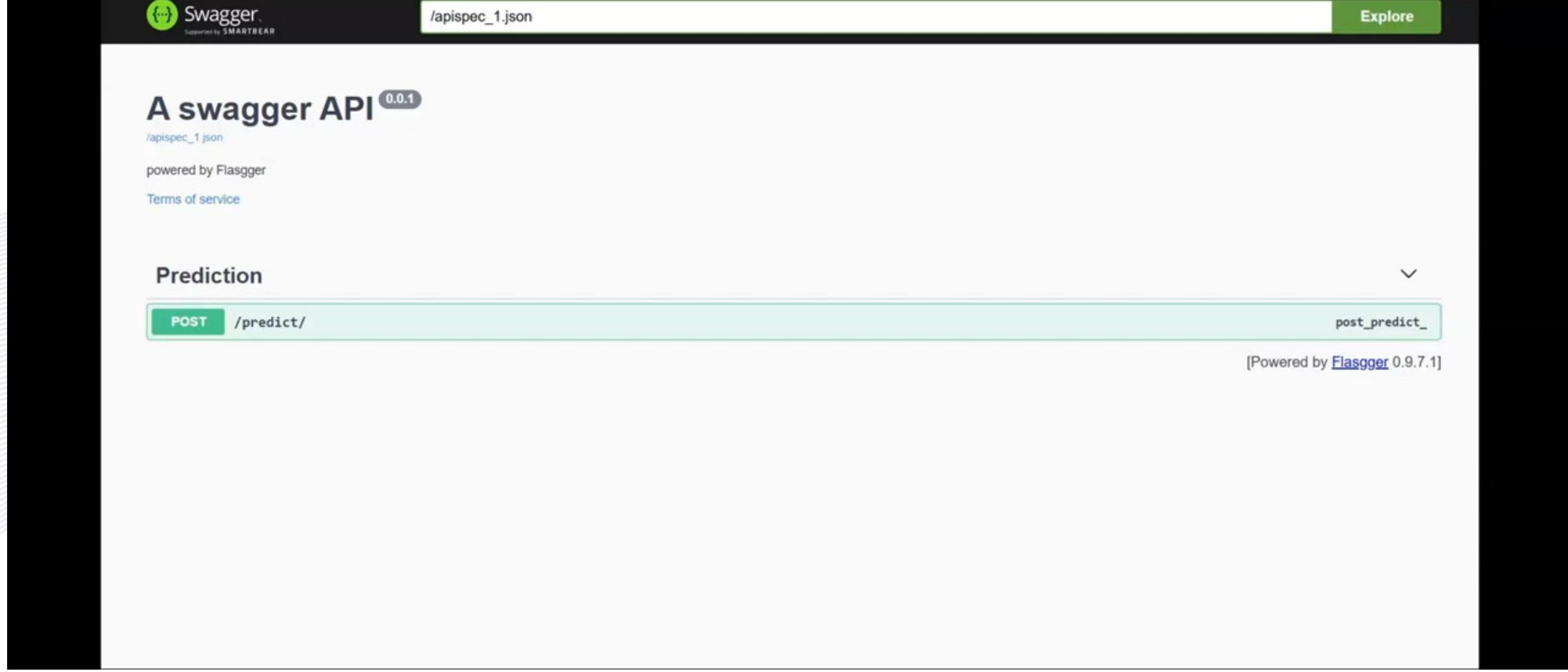
The **ISIS** ORG has claimed responsibility for a suicide bomb blast in the **Tunisian** LOC capital **earlier this week** DATE, the **militant group** ORG 's **Amaq news agency** ORG said on **Thursday** DATE. A **militant** PER wearing an **explosives belt** WEAPON blew himself up in **Tunis** LOC

PROJEMİZİN YOL HARİTASI

Model performansının iyileştirilmesine yönelik çalışmalar yapılabilir



DEMO VIDEO



Videomuzun Youtube Linki: <https://www.youtube.com/watch?v=P4aSHKNMZ5E>

Videomuzun Drive Linki: <https://drive.google.com/drive/folders/1Ij2oH1FCjSBEEYY0BhkMGtptB8QdWdO36>

TEŞEKKÜRLER