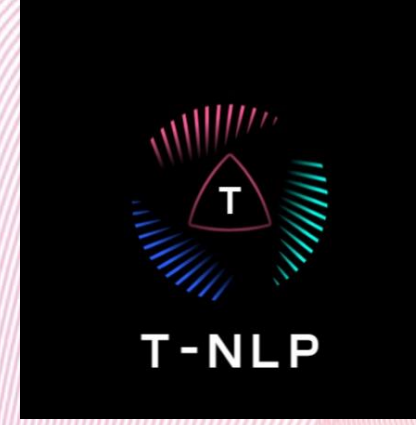


Doğal Dil İşleme Senaryosu



Türkçe Doğal Dil İşleme

8 - 9 Ağustos 2024

T-NLP

EKİBİMİZ



Güüzade Evni

- Projede uygun parametrelerin seçimi, arayüz geliştirme ve modelleme aşamalarında görev almaktadır.



Zeynep Baydemir

- Projede veri önişleme, veri toplama ve model aşamalarında görev almaktadır.



Kübra Arslan

- Projede kaynak belirleme, veri önişleme, veri toplama ve arayüz geliştirme aşamalarında görev almaktadır.

PROJENİN TANIMI

Bu yarışma senaryosu, "Entity Bazlı Duygu Analizi" üzerine odaklanmaktadır.

Katılımcılardan, müşteri yorumlarındaki belirli hizmet veya ürünlerle ilgili entity'leri doğru bir şekilde tanımlamaları ve bu entity'lerle ilgili duyguları (olumlu, olumsuz, nötr) sınıflandırmaları beklenmektedir. Yarışma, yorumları doğru entity'lere atfetme ve hizmet/ürün özelliklerine göre duygu analizi yapma becerilerini test eder.

```
{...} output
1  {
2    ▼ "entity_list": [
3      "0": "turkcell"
4    ],
5    ▼ "results": [
6      ▼ "0": {
7        "entity": "turkcell",
8        "sentiment": "olumsuz"
9      }
10   ]
11 }
```


PROJENİN SAĞLADIĞI ÇÖZÜM

Çalışmanın Sağladığı Çözüm:

Bu çalışma, müşteri geri bildirimlerindeki belirli entity'lere (şirketler, ürünler, hizmetler vb.) yönelik duyguları doğru bir şekilde sınıflandırmak için gelişmiş doğal dil işleme (NLP) modelleri kullanarak bir çözüm sunar. Kullanılan üç model (BERT, DistilBERT, Electra), müşteri yorumlarını analiz ederek, yorumlarda bahsedilen entity'leri tanımlar ve bu entity'lere yönelik duyguları (olumlu, olumsuz, nötr) sınıflandırır.

Bu çözüm, geniş ölçekli müşteri geri bildirimlerini otomatik olarak analiz edebilir ve firmaların ürün ve hizmet kalitesini anlamalarına yardımcı olur. Modelin sağladığı detaylı analizler, firmaların hangi alanlarda güçlü olduğunu ve hangi alanlarda iyileştirmeler yapmaları gerektiğini belirlemelerini sağlar.

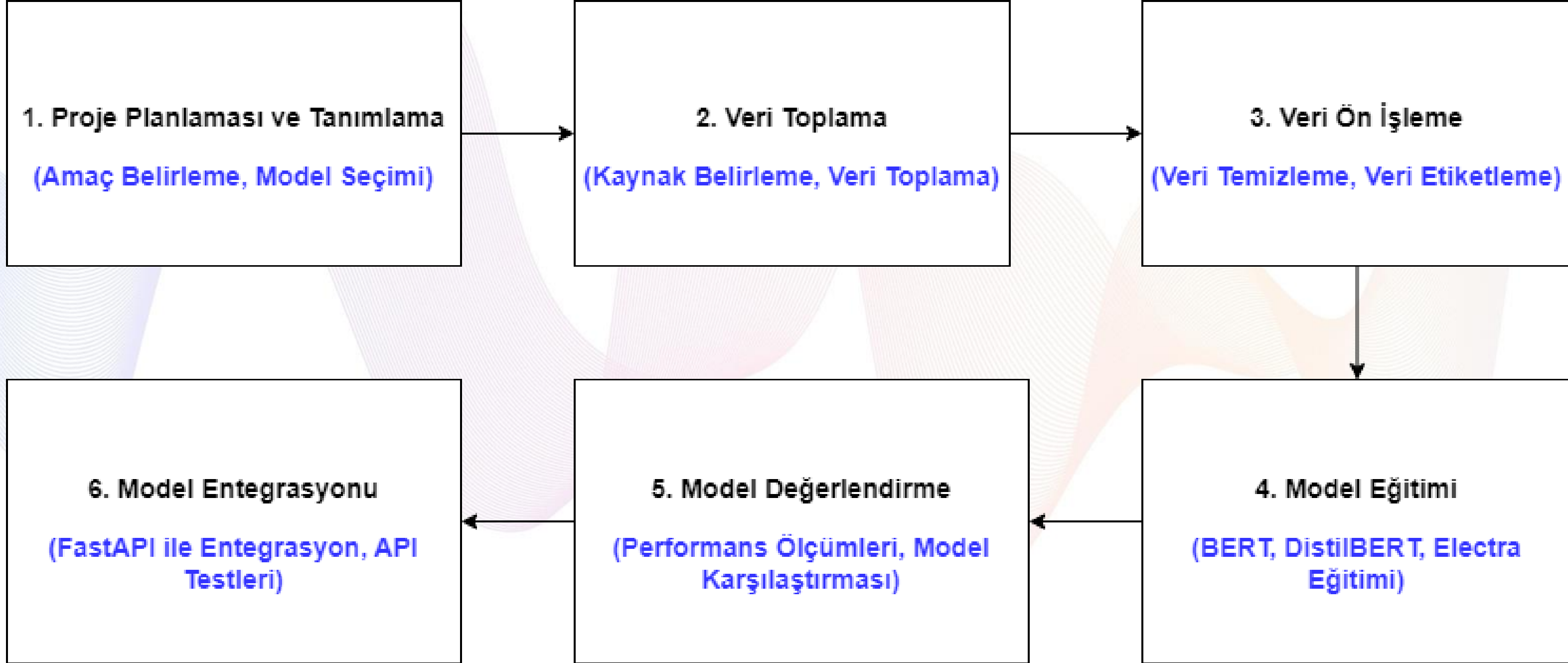
PROJENİN SAĞLADIĞI ÇÖZÜM

Hedef Kitle:

- Şirketler ve Kurumlar
- Pazarlama ve Müşteri İlişkileri Ekipleri
- Veri Bilimciler ve NLP Uzmanları
- Araştırmacılar ve Akademisyenler



PROJE İŞ AKIŞI



VERİ SETİ

- Çalışma boyunca toplam **50.845** veri çekilmiştir.
- Veriler, Ekşi Sözlük, X ve Şikayetvar platformlarından çekilmiştir.
- Her bir veri seti, farklı yöntemlerle toplanmış ve ayrı dosyalar halinde kaydedilmiştir.

The logo for Ekşi Sözlük, featuring a green teardrop icon followed by the text "ekşisözlük" in a green, lowercase, sans-serif font.The logo for Şikayetvar, featuring a green checkmark icon followed by the text "şikayetvar" in a dark blue, lowercase, sans-serif font.

VERİ SETİ

Ekşi Sözlük Verileri: Ekşi Sözlük'teki kullanıcı yorumları, belirli başlıklar altında toplanmıştır. Bu işlem için Selenium ve BeautifulSoup kütüphaneleri kullanılmıştır.

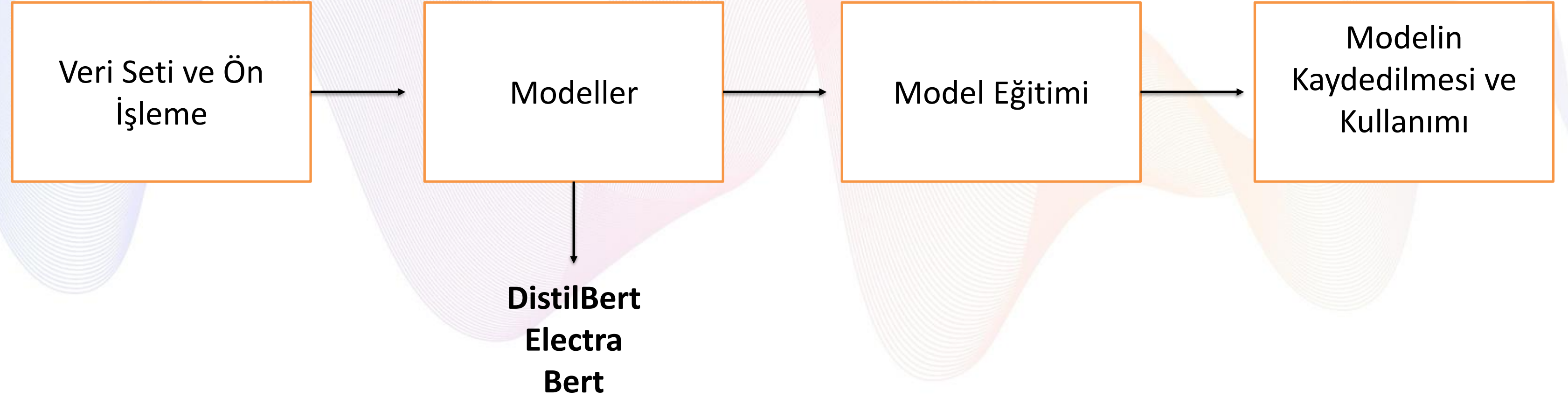
X Verileri: Veriler, X (eski adıyla Twitter) uygulamasında Turkcell'e ait olan Superonline ve Turkcell Müşteri Hizmetleri isimli X hesaplarından alınmıştır.

Selenium kullanılarak, sayfanın kaydırılması ve tweetlerin yakalanması sağlanmıştır.

Şikayetvar Verisi: Şikayetvar platformundaki şikayetler belirlenen başlığa göre toplanarak analiz için kullanılmıştır.



YÖNTEM VE TEKNİKLER



MODEL EĞİTİMİ VE DEĞERLENDİRME



Eksik ve yinelenen veriler temizlenmiş, veri seti analiz için hazır hale getirilmiştir.



Model Seçimi
BERT, Electra, DistilBert modelleri kullanılmıştır.



Metinler, belirlenen maksimum token uzunluğunda olacak şekilde token'lara dönüştürülmüştür.

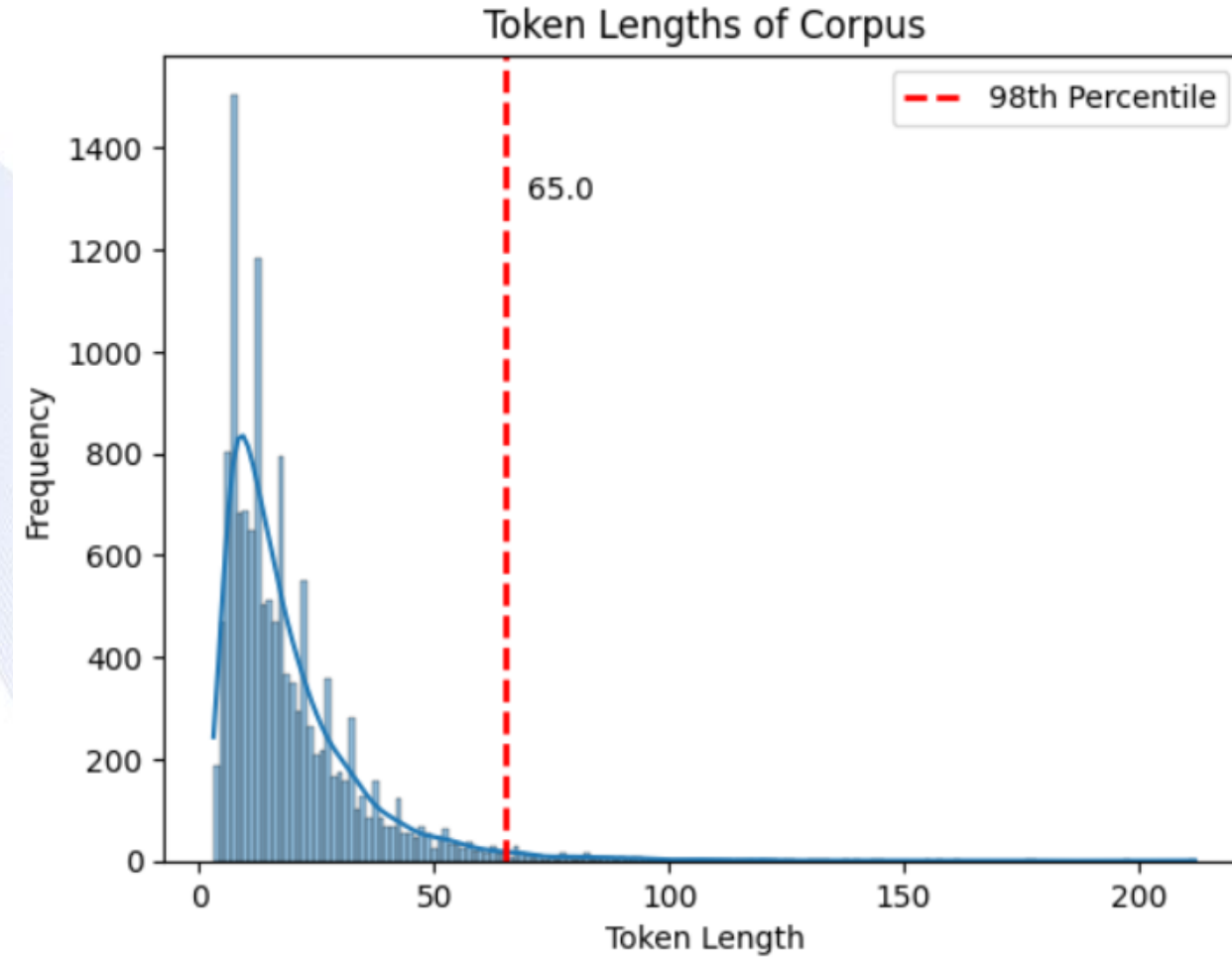


Token'lar PyTorch tensörlerine dönüştürülmüş ve %90 eğitim, %10 doğrulama seti olarak ayrılmıştır.



AdamW optimizatörü ve öğrenme oranı zamanlayıcısı kullanılarak modelin eğitim süreci iyileştirilmiştir. Model, belirlenen epoch değerleri boyunca eğitilmiş ve her epoch sonunda eğitim kaybı ile doğrulama performansı değerlendirilmiştir.

SONUÇLAR



	Validation accuracy	Validation loss	Training loss
Bert	0.9144	0.2378	0.1516
DistilBert	0.8996	0.2771	0.2488
Electra	0.9189	0.2378	0.1497

PROJE YOL HARİTASI

Proje, gelecekteki çalışmalarda daha fazla veri kaynağı ve gelişmiş NLP teknikleriyle genişletilerek, analiz derinliği artırılabilir. Ayrıca, gerçek zamanlı duygu analizi ve otomatik uyarı sistemleri eklenerek proaktif çözümler sunan bir yapıya dönüştürülebilir. Veri gizliliği ve etik konularındaki çalışmalar da projenin güvenilirliğini güçlendirecektir.



DEMO VIDEO

<https://youtu.be/1DegOrx07wE>

TEŞEKKÜRLER