



**SOSYAL MEDYA İZLEME VE DEPREM ACİL  
DURUMLARINA HIZLI YANIT: TWITTER VERİSİ  
ÜZERİNE BİR ÇALIŞMA**

**Dilara CANDAN KARADAŞ**

**Danışman: Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER**  
**Yüksek Lisans Tezi**

**Bilgisayar Mühendisliği Ana Bilim Dalı**

**2023**

(Her hakkı saklıdır.)

T.C.  
ATATÜRK ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI

**SOSYAL MEDYA İZLEME VE DEPREM ACİL DURUMLARINA HIZLI YANIT:  
TWITTER VERİSİ ÜZERİNE BİR ÇALIŞMA**

(Social Media Monitoring and Rapid Response To Earthquake Emergencies: A Study on  
Twitter Data)

YÜKSEK LİSANS TEZİ

Dilara CANDAN KARADAŞ

Dr. Öğr. Üyesi Gülşah TUMÜKLÜ ÖZYER

Erzurum  
Aralık, 2023

## KABUL VE ONAY TUTANAĞI

Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER danışmanlığında, Dilara Candan Karadaş tarafından hazırlanan bu çalışma, 08/12/2023 tarihinde aşağıdaki jüri tarafından Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Bilim Dalı'nda Yüksek Lisans tezi olarak **oybirliği** ile kabul edilmiştir.

Jüri Başkanı: Doç. Dr. Alev MUTLU

*Kocaeli Üniversitesi*

Aslı ıslak imzalıdır

Danışman: Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER

*Atatürk Üniversitesi*

Aslı ıslak imzalıdır

Jüri Üyesi: Doç. Dr. Önder ÇOBAN

*Atatürk Üniversitesi*

Aslı ıslak imzalıdır

Enstitü Yönetim  
Kurulunun .../.../.... tarih  
ve ..... sayılı kararı.

Bu tezin Atatürk Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili maddelerinde belirtilen şartları yerine getirdiğini onaylarım.

**Prof. Dr. Saltuk Buğrahan CEYHUN**

**Enstitü Müdürü**

Aslı ıslak imzalıdır

**Not:** Bu tezde kullanılan özgün ve başka kaynaklardan yapılan bildiriş, çizelge, şekil ve fotoğrafların kaynak olarak kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

## ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU

Yüksek Lisans Tezi olarak Dr. Öğr. Üyesi Gülşah Tümüklü Özyer danışmanlığında sunulan “ Sosyal Medya İzleme ve Deprem Acil Durumlarına Hızlı Yanıt: Twitter Verisi Üzerine Bir Çalışma” başlıklı çalışmanın tarafımızdan bilimsel etik ilkelere uyularak yazıldığını, yararlanılan eserlerin kaynakçada gösterildiğini, Fen Bilimleri Enstitüsü tarafından belirlenmiş olan Turnitin Programı benzerlik oranlarının aşılmadığını ve aşağıdaki oranlarda olduğunu beyan ederiz.

| Tez Bölümleri                   | Tezin Benzerlik Oranı (%) | Maksimum Oran (%) |
|---------------------------------|---------------------------|-------------------|
| Giriş                           | 0                         | 30                |
| Kuramsal Temeller               | 0                         | 30                |
| Materyal ve Metot               | 0                         | 35                |
| Araştırma Bulguları ve Tartışma | 0                         | 20                |
| Sonuçlar ve Öneriler            | 0                         | 20                |
| Tezin Geneli                    | 0                         | 25                |

**Not:** Yedi kelimeye kadar benzerlikler ile Başlık, Kaynakça, İçindekiler, Teşekkür, Dizin ve Ekler kısımları tarama dışı bırakılabilir. Yukarıdaki azami benzerlik oranları yanında tek bir kaynaktan olan benzerlik oranlarının %5'ten büyük olmaması gerekir.

Beyan edilen bilgilerin doğru olduğunu, aksi halde doğacak hukuki sorumlulukları kabul ettiğimizi beyan ederiz.

|                            |                                     |
|----------------------------|-------------------------------------|
| Tez Yazarı (Öğrenci)       | Tez Danışmanı                       |
| Dilara CANDAN KARADAŞ      | Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER |
| 08.12.2023                 | 08.12.2023                          |
| İmza: Aslı ıslak imzalıdır | İmza: Aslı ıslak imzalıdır          |

\* Tez ile ilgili YÖKTEZ’de yayınlamasına ilişkin bir engelleme var ise aşağıdaki alanı doldurunuz.

☐ Tezle ilgili patent başvurusu yapılması / patent alma sürecinin devam etmesi sebebiyle Enstitü Yönetim Kurulunun ....../....../.... tarih ve ..... sayılı kararı ile teze erişim 2 (iki) yıl süreyle engellenmiştir.

☐ Enstitü Yönetim Kurulunun ....../....../.... tarih ve ..... sayılı kararı ile teze erişim 6 (altı) ay süreyle engellenmiştir.

## TEŞEKKÜR

Bu tez çalışmasının hazırlık sürecindeki sabır, destek ve ilgileri için öncelikle değerli danışman hocama Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER teşekkürlerimi sunarım.

Hiçbir zaman benden desteklerini esirgemeyen ve hayatımın her döneminde yanımda olan değerli aileme şükranlarımı sunuyorum.

Bu çalışmanın tamamlanmasında bana sürekli destek olan ve sürecin her aşamasında yanımda bulunan eşim Melih KARADAŞ'a derin minnettarlığımı sunarım.

Tez çalışma sürecimdeki katkılarından ötürü Özge Nur ÖZARAS'a içten teşekkürlerimi sunarım.

İyi ki varsınız.

Dilara CANDAN KARADAŞ

## ÖZET

### YÜKSEK LİSANS TEZİ

#### SOSYAL MEDYA İZLEME VE DEPREM ACİL DURUMLARINA HIZLI YANIT: TWITTER VERİSİ ÜZERİNE BİR ÇALIŞMA

Dilara CANDAN KARADAŞ

Danışman: Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER

**Amaç:** Deprem, ülkemizin karşı karşıya olduğu en kaçınılmaz doğal afetlerden biridir. Tarihsel süreçte pek çok kez büyük kayıplara yol açan bu doğal olayın getirdiği tehlikelerin yönetilmesi, toplumun can ve mal güvenliği için son derece kritiktir. Depremi hemen ardından alınması gereken hızlı aksiyonlar için anlık ve doğru veriye erişim büyük önem taşımaktadır. Ancak, bu kritik dönemde toplanan veri genellikle yoğun ve karmaşıktır, bu nedenle sadece sınırlı miktarda veri etiketlemesi mümkündür. Bu tez çalışmasının öncelikli amacı, böyle bir durumda Twitter'da meydana gelen veri kirliliği içerisinde, afet yönetimi ve müdahale için kritik öneme sahip bilgileri hızlı ve etkin bir şekilde sınıflandırabilmektir. Bu sürecin en önemli zorluğu, az sayıda etiketli veri ile maksimum verimliliğe ulaşmayı hedeflemektir. Bu bağlamda, tez çalışmasında transfer öğrenme modellerinin yanı sıra, geleneksel makine öğrenmesi yöntemleri ve sıfır atış (zero-shot) ile birkaç örnek öğrenme (few-shot) yöntemleri incelenmiş, ek olarak farklı modeller de ele alınarak deprem sonrası sosyal medya verisinin sınıflandırılmasında modellerin etkinliğini değerlendirmeyi amaçlamaktadır.

**Yöntem:** Bu çalışmada Türkçe Twitter verileri ile metin sınıflandırma amacıyla ülkemizde en son meydana gelen ve çok büyük yıkımlara sebep olan 6 Şubat 2023 depremine ait Twitter mesajları üzerinde çalışılmıştır. Bu depreme ait hali hazırda derlenmiş bir tweet veri seti araştırılarak bulunmuştur (Kaggle Turkey and Syria Earthquake Tweets). Bu veri setinden 2000 tane veri seçilerek 3 sınıf olacak şekilde etiketleme yapılmıştır. Bu sınıflar ilgisiz (0), enkaz (1) ve ihtiyaç (2) şeklindedir. Elde bulunan az sayıda etiketli veri ile transfer öğrenme yöntemlerinden olan BERT, RoBERTa ve DistilBERT kullanılarak eğitimler yapılmıştır. Bu modeller ile sınırlı kalmayıp geleneksel yöntemler ile de değerlendirilerek kıyaslama yapılmıştır. Tüm bunlara ek olarak daha az veri ile de deneyler yapılarak sıfır atış (zero-shot) ile birkaç örnek öğrenme (few-shot) yöntemleri de çalışmaya dahil edilmiştir. Bu eğitimlerde alınan sonuçlara göre en başarılı sonuçları veren model belirlenmiştir.

**Bulgular:** Çalışma kapsamında 6 Şubat Depremi ile 3 etiketli yeni bir veri seti oluşturulmuştur. Az sayıda etiketli veri ile başarılı sonuçları elde etmek hedeflenmiştir. Alınan sonuçlar detaylı olarak incelendiğinde kullanılan modellerin beklediği gibi sonuçlar verdiği gözlemlenmiştir. Farklı veri setleri de kullanılarak sonuçlar üzerindeki etkisi incelenmiştir. Sonuçlara göre hazırlanan veri setinin tamamını kullanmadan bile çok az bir kısmını kullanarak da benzer sonuçlar elde edilmiştir.

**Sonuç:** Bu çalışma kapsamında bir afet sonrasında sosyal medyadan gelen verileri sınıflandırarak anlamlı hale getirebilmek hedeflenmiştir. Sonuç olarak seçilen yöntemler ile yapılan çalışmaların sonucunda bu amaca uygun olduğu gözlemlenmiştir.

**Anahtar Kelimeler:** Makine Öğrenmesi, Transfer Öğrenme, Metin Sınıflandırma, Doğal Dil İşleme, sıfır atış (zero-shot), birkaç örnek öğrenme (few-shot), Twitter, Deprem

Aralık 2023, 72 sayfa

## ABSTRACT

### MASTER THESIS

#### SOCIAL MEDIA MONITORING AND RAPID RESPONSE TO EARTHQUAKE EMERGENCIES: A STUDY ON TWITTER DATA

Dilara CANDAN KARADAŞ

Supervisor: Assist. Prof. Dr. Gülşah TÜMÜKLÜ ÖZYER

**Purpose:** Earthquakes are among the most inevitable natural disasters our country faces. Throughout history, this natural phenomenon has often led to significant losses, making the management of its dangers critical for the safety and security of the public. Access to instant and accurate data is crucial for swift actions following an earthquake. However, the data collected during this critical period is often dense and complex, limiting the amount of data that can be effectively labeled. The primary aim of this thesis is to quickly and effectively classify information crucial for disaster management and response amid the data pollution on Twitter during such events. The main challenge of this process is to achieve maximum efficiency with a limited amount of labeled data. In this context, the thesis examines not only transfer learning models but also traditional machine learning methods, zero-shot, and few-shot learning approaches. Additionally, it aims to evaluate the effectiveness of different models in classifying post-earthquake social media data.

**Method:** In this study, Turkish Twitter data related to the most recent devastating earthquake that struck our country on February 6, 2023, has been analyzed for text classification purposes. A dataset of tweets regarding this earthquake has been compiled and located (Kaggle Turkey and Syria Earthquake Tweets). From this dataset, 2000 data points have been selected and labeled into 3 classes: irrelevant (0), debris (1), and needs (2). Training has been conducted using transfer learning methods such as BERT, RoBERTa, and DistilBERT with this limited labeled data. Not confined to these models, traditional methods have also been evaluated for comparison. Additionally, less data-intensive approaches such as zero-shot and few-shot learning methods have been incorporated into the study. Based on the results of these trainings, the most successful model has been determined.

**Findings:** The study involved creating a new dataset labeled with 3 tags associated with the February 6th Earthquake. The aim was to achieve successful results with a limited number of labeled data. A detailed examination of the results revealed that the models used performed as expected. The impact of using different datasets was also analyzed. According to the findings, similar results were obtained even without using the entire dataset, but just a small portion of it.

**Conclusion:** The objective of this study was to classify data from social media after a disaster, making it meaningful. As a result, it has been observed that the methods chosen for the studies conducted are suitable for this purpose.

**Keywords:** Machine Learning, Transfer Learning, Text Classification, Natural Language Processing, Zero-Shot Learning, Few-Shot Learning, Twitter, Earthquake.

December 2023, 72 pages

## İÇİNDEKİLER

|                                                            |      |
|------------------------------------------------------------|------|
| KABUL VE ONAY TUTANAĞI.....                                | i    |
| ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU .....                 | ii   |
| TEŞEKKÜR .....                                             | iii  |
| ÖZET .....                                                 | iv   |
| ABSTRACT .....                                             | v    |
| TABLolar DİZİNİ.....                                       | viii |
| ŞEKİLLER DİZİNİ .....                                      | ix   |
| KISALTMALAR VE SİMGELER DİZİNİ .....                       | xi   |
| GİRİŞ.....                                                 | 1    |
| Tezin Katkıları .....                                      | 3    |
| KURAMSAL TEMELLER.....                                     | 5    |
| Doğal Dil İşleme .....                                     | 7    |
| Metin Sınıflandırma .....                                  | 7    |
| Veri Seti Toplama .....                                    | 9    |
| Veri Etiketleme .....                                      | 10   |
| Veri Ön İşleme .....                                       | 11   |
| Geleneksel Makine Öğrenmesi .....                          | 12   |
| Transfer Öğrenme Yöntemleri .....                          | 15   |
| BERT modeli .....                                          | 16   |
| RoBERTa modeli .....                                       | 18   |
| DistilBERT modeli .....                                    | 19   |
| Zero-Shot (Sıfır Atış) öğrenme yöntemi .....               | 19   |
| Few-Shot (Birkaç Atış) öğrenme yöntemi .....               | 20   |
| TARS modeli .....                                          | 20   |
| Literatür Taraması .....                                   | 22   |
| MATERYAL VE METOT.....                                     | 26   |
| Veri Seti .....                                            | 26   |
| Yöntem.....                                                | 31   |
| Geleneksel makine öğrenmesi yöntemlerinin uygulanması..... | 32   |
| Transfer öğrenme yöntemlerinin uygulanması .....           | 33   |

|                                                                                                  |    |
|--------------------------------------------------------------------------------------------------|----|
| Sıfır atış öğrenme (Zero-Shot) - birkaç örnek öğrenme (Few-Shot) yöntemlerinin uygulanması ..... | 34 |
| ARAŞTIRMA BULGULARI VE TARTIŞMA.....                                                             | 37 |
| Geleneksel Yöntemler ile Sınıflandırma Sonuçları.....                                            | 37 |
| Transfer Öğrenme Yöntemleri ile Sınıflandırma Sonuçları.....                                     | 41 |
| Sıfır Atış ve Birkaç Örnek Öğrenme ile Sınıflandırma Sonuçları .....                             | 44 |
| BERT(ft) için sınıflandırma sonuçları .....                                                      | 44 |
| TARS için sınıflandırma sonuçları .....                                                          | 49 |
| SONUÇ VE ÖNERİLER .....                                                                          | 55 |
| KAYNAKÇA .....                                                                                   | 56 |
| ÖZGEÇMİŞ.....                                                                                    | 59 |



## TABLÖLAR DİZİNİ

|                                                                                                                    |    |
|--------------------------------------------------------------------------------------------------------------------|----|
| <b>Tablo 1.</b> Örnek Bir Tweet Üzerinde Ön İşleme .....                                                           | 12 |
| <b>Tablo 2.</b> Veri Setindeki Twitter Mesajlarına Ait Bilgiler .....                                              | 26 |
| <b>Tablo 3.</b> Veri Setinin Dillere Göre Dağılımı .....                                                           | 27 |
| <b>Tablo 4.</b> Seçilen Etiketler ve Açıklaması.....                                                               | 28 |
| <b>Tablo 5.</b> Örneklerin Kategoriler Arası Dağılımı.....                                                         | 28 |
| <b>Tablo 6.</b> Geleneksel Yöntemler ile Sınıflandırma Sonuçları .....                                             | 37 |
| <b>Tablo 7.</b> Geleneksel Yöntemler ile Sınıflandırma Sonuçları .....                                             | 39 |
| <b>Tablo 8.</b> BERT, RoBERTa ve DistilBERT Sonuçları .....                                                        | 41 |
| <b>Tablo 9.</b> BERT, RoBERTa ve DistilBERT İçin Çapraz Doğrulama Sonuçları.....                                   | 44 |
| <b>Tablo 10.</b> BERT(ft) Sıfır Atış ve Birkaç Örnek ile Öğrenme Sonuçları .....                                   | 44 |
| <b>Tablo 11.</b> BERT(ft) Farklı Veri Seti ile Eğitim Sonrası Birkaç Örnek Öğrenme Yöntemi<br>Sonuçları .....      | 46 |
| <b>Tablo 12.</b> BERT(Ft) Farklı Konuda Veri Seti ile Eğitim Sonrası Birkaç Örnek Öğrenme<br>Sonuçları .....       | 48 |
| <b>Tablo 13.</b> TARS Modeli Benzer Veri Seti ile Eğitim Sonrası Birkaç Örnek ile Öğrenme<br>Sonuçları .....       | 49 |
| <b>Tablo 14.</b> TARS Modelinde Farklı Konuda Veri Seti ile Eğitim Sonrası Birkaç Örnek<br>Öğrenme Sonuçları ..... | 51 |
| <b>Tablo 15.</b> TARS Çapraz Doğrulama Sonucu.....                                                                 | 53 |

## ŞEKİLLER DİZİNİ

|                                                                                                                           |    |
|---------------------------------------------------------------------------------------------------------------------------|----|
| Şekil 1. Tek etiketli ve çok etiketli metin sınıflandırma örneği .....                                                    | 8  |
| Şekil 2. Denetimli ve denetimsiz metin sınıflandırma .....                                                                | 14 |
| Şekil 3. Geleneksel makine öğrenmesi ve transfer öğrenme.....                                                             | 15 |
| Şekil 4. Maskelenmiş dil modeli ve sonraki cümle tahmini .....                                                            | 18 |
| Şekil 5. TARS modeli mimarisi .....                                                                                       | 22 |
| Şekil 6. Twitter API'den twitter mesajları alınırken kullanılan hashtag dağılımı.....                                     | 27 |
| Şekil 7. Enkaz olarak etiketli twitter mesajlarının ön işleme öncesi durumu .....                                         | 29 |
| Şekil 8. Enkaz olarak etiketli twitter mesajlarının ön işleme sonrası durumu .....                                        | 29 |
| Şekil 9. İhtiyaç olarak etiketli twitter mesajlarının ön işleme öncesi durumu .....                                       | 29 |
| Şekil 10. İhtiyaç olarak etiketli twitter mesajlarının ön işleme sonrası durumu .....                                     | 30 |
| Şekil 11. İlgisiz olarak etiketli twitter mesajlarının ön işleme öncesi durumu .....                                      | 30 |
| Şekil 12. İlgisiz olarak etiketli twitter mesajlarının ön işleme sonrası durumu .....                                     | 30 |
| Şekil 13. 6 Şubat deprem verisi kelime bulutu.....                                                                        | 31 |
| Şekil 14. Ürün yorumları içeren veri seti kelime bulutu .....                                                             | 31 |
| Şekil 15. CrisisNLP veri seti kelime bulutu .....                                                                         | 31 |
| Şekil 16. Geleneksel yöntemlere ait karmaşıklık matrisi .....                                                             | 38 |
| Şekil 17. Optimal K değeri için grafik .....                                                                              | 39 |
| Şekil 18. Geleneksel yöntemlere ait karmaşıklık matrisi .....                                                             | 40 |
| Şekil 19. Optimal K değeri için grafik .....                                                                              | 41 |
| Şekil 20. BERT eğitim loss grafiği .....                                                                                  | 42 |
| Şekil 21. RoBERTa eğitim loss grafiği .....                                                                               | 42 |
| Şekil 22. DistilBERT eğitim loss grafiği .....                                                                            | 42 |
| Şekil 23. BERT, RoBERTa ve DistilBERT için karmaşıklık matrisleri.....                                                    | 43 |
| Şekil 24. BERT(ft) sıfır atış ve birkaç örnek öğrenme karmaşıklık matrisi .....                                           | 45 |
| Şekil 25. BERT(ft) farklı veri seti ile eğitim sonrası birkaç örnek öğrenme yöntemi ile<br>karmaşıklık matrisi .....      | 47 |
| Şekil 26. BERT(ft) Farklı konuda veri seti ile eğitim sonrası birkaç örnek öğrenme<br>karmaşıklık matrisi .....           | 48 |
| Şekil 27. TARS modeli benzer veri seti ile eğitim sonrası birkaç örnek ile öğrenme<br>karmaşıklık matrisi sonuçları ..... | 50 |

|                                                                                                                                      |    |
|--------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Şekil 28.</b> TARS modeli farklı konulu veri seti ile eğitim sonrası birkaç örnek ile öğrenme karmaşıklık matrisi sonuçları ..... | 52 |
|--------------------------------------------------------------------------------------------------------------------------------------|----|



## KISALTMALAR VE SİMGELER DİZİNİ

|                   |                                                                                     |
|-------------------|-------------------------------------------------------------------------------------|
| <b>API</b>        | : Application Programming Interface (Uygulama Programlama Arayüzü)                  |
| <b>BERT</b>       | : Bidirectional Encoder Representations from Transformers (Çift Yönlü)              |
| <b>DistilBERT</b> | : Distillation BERT (Damıtılmış BERT)                                               |
| <b>KNN</b>        | : K-Nearest Neighbours (K En Yakın Komşu Kodlayıcı Gösterimi)                       |
| <b>MLM</b>        | : Masked Language Model (Maskelenmiş Dil Modeli)                                    |
| <b>NLP</b>        | : Natural Language Processing (Doğal Dil İşleme)                                    |
| <b>NSP</b>        | : Next Sentence Prediction (Sonraki Cümle Tahmini)                                  |
| <b>RoBERTa</b>    | : Robustly Optimized BERT (Güçlü Bir Şekilde Optimize Edilmiş BERT)                 |
| <b>ROC</b>        | : Alıcı İşletim Özelliği (Receiver Operating Characteristic)                        |
| <b>SVM</b>        | : Destek Vektör Makinesi (Support Vector Machines)                                  |
| <b>TARS</b>       | : Task-Aware Representation of Sentences (Cümlelerin Göreve Duyarlı Temsili)        |
| <b>Tf-Idf</b>     | : Term Frequency -Inverse Document Frequency (Terim Frekansı-Ters Doküman Frekansı) |
| <b>URL</b>        | : Uniform Resource Locator (Birleşik Kaynak Konumlandırıcısı)                       |

## GİRİŞ

İnternet ve sosyal medya teknolojilerinin yaygınlaşmasıyla birlikte, çeşitli konulardaki istek ve beklentiler, kullanıcılar tarafından bu mecralar üzerinden ifade edilmeye başlanmıştır. Özellikle savaşlar, depremler, salgınlar gibi olağanüstü durumlar karşısında, sosyal medya kullanıcılarının zaman ve mekân sınırlamaları olmaksızın istek, duygu ve düşüncelerini iletmelerine imkân tanımıştır. Doğal afetler sonucunda iletişim altyapısının tahrip olması durumlarında, sosyal medyanın, zaman ve mekân kavramlarının sınırlarını aşarak haber ve bilgilerin geniş kitlelere anında ulaşmasına olanak sağladığı görülmüştür. Ayrıca, sosyal medya aracılığıyla bireylerin yardım talepleri, dayanışma çağrıları ve düşüncelerin ifade edilmesine de imkân verilmiş, doğal afetler karşısında alınan veya alınmayan tedbirlere yönelik çeşitli tepkilerin gösterilmesinde bir araç olmuştur.

Genel bir ifadeyle deprem, yangın, sel, kuraklık, tsunami, çığ gibi doğal afetler sonucunda sosyal medya gerek afetzedelerin yardım ve isteklerini geniş kitlelere iletmesi gerekse de içinde bulundukları duruma çözüm arayışları gerekse de sorumlu olan yetkililere tepki gösterilmesi açısından birçok özelliği içerisinde barındırmaktadır. Dolayısıyla son yıllarda afetler konusunda sosyal medyanın rolünün kaçınılmaz olduğu görülmektedir (Argın, Y., 2023). Sosyal medyanın krizlere yanıt olarak giderek daha popüler bir haber kaynağı haline geldiği gözlemlenmektedir (Alam & Imran, 2020). Sosyal medyanın, krizlere yanıt ve toplulukların iyileşme süreçlerindeki artan rolü, birçok araştırmacının dikkatini çekmektedir (Zhang *et al.* 2019). Özellikle, sosyal medyanın, yardım operasyonlarını organize etme ve topluluk desteğini mobilize etme potansiyeline de vurgu yapılmaktadır (Meier, 2015).

Günümüz bilgi çağında, özellikle sosyal medyanın yaygın olarak kullanıldığı bu dönemde, gerçek zamanlı verilere dayanarak alınan bilgilerin hızlı ve etkin bir şekilde değerlendirilmesi büyük bir öneme sahiptir (Nguyen *et al.* 2017). Bu bilgilerin büyük çoğunluğu sosyal medya platformları üzerinden gelmektedir. Sosyal medya üzerinden gelen bilgiler genellikle yüksek hacimde ve oldukça hızlı olduğu görülmektedir. Dolayısıyla bu durum, bilgilerin manuel olarak işlenmesini zorlaştırmaktadır. Bu, doğru ve anlamlı bilgileri otomatik olarak filtrelemenin ve tanımlamanın kritik bir ihtiyaç haline geldiği anlamına gelmektedir.

Twitter gibi sosyal medya platformları, topluluklar için gerçek zamanlı bilgi paylaşımı sağlamaktadır ve kriz anlarında kritik bir rol oynadığı görülmüştür. Örneğin, 2019'da Notre

Dame yangını sırasında 17.000'den fazla tweet gönderilmişti. Benzer şekilde, 2023'teki Türkiye ve Suriye depreminde, pek çok kişinin yardım çağrıları ve durum güncellemeleri Twitter üzerinden paylaşılmıştır (Toraman *et al.* 2023). Bu çağrılar birçok kişinin kurtulmasında veya gönderilen yardımlara ulaşmasında oldukça büyük öneme sahiptir

Birçok kurum ve kuruluş, olaylara ilişkin güncel bilgilere ulaşmak için sosyal medya platformlarını incelemektedir. Bu platformlardaki yoğun veri akışını el ile işlemek oldukça güçtür ve bu tür görevler için gelişmiş teknolojiye sahip yazılımlar tercih edilmektedir. Bu yazılımlar, insanların yerine geçerek, veriyi ele alıp, analiz edip, sınıflandırarak ve önemli bilgileri çıkararak işler. Bu süreçlerde yapay zekâ teknolojilerinden, özellikle de makine öğrenmesi ve doğal dil işleme gibi alanlardan faydalanılmaktadır.

Doğal dil işleme (NLP), insanların kullandığı dilin bilgisayarlar tarafından yorumlanması ve anlamlandırılması üzerine yoğunlaşan bir bilim dalıdır. Birçok alt problemi içerisinde barındıran bu alanda, metinleri belirli kategorilere ayırma işlemi olan metin sınıflandırma da büyük bir öneme sahiptir. Metin sınıflandırma, belgeleri veya metinleri, özellikle sosyal medya metinleri gibi büyük ve eşit dağılımı olmayan veri kaynaklarını, önceden tanımlanmış kategorilere ayırmak için kullanılan bir süreçtir. Bu otomatik sınıflandırma işlemi, kullanıcı yorumları, Twitter mesajları veya blog gönderileri gibi metinleri analiz ederek belirli temaları veya duygusal tonları tespit etmek için özellikle yararlıdır. Ancak, dillerin karmaşık yapıları metin sınıflandırmanın çeşitli zorluklarla karşılaşmasına yol açmaktadır (Franko 2018). Dilbilimsel yapısal farklılıklar ve özellikler, metin analizi alanında çalışmayı zorlaştıran faktörler arasında yer almaktadır. Metinlerin analiz sürecinde, harf, kelime, cümle veya tam belge bazında olmak üzere farklı seviyelerde yaklaşımlar uygulanabilmektedir. Bu yaklaşımlar, analizin yapısını ve derinliğini belirlemekte, metinlerin detaylı incelenmesine olanak tanımaktadır. Bu nedenle, doğal dil işlemeyle metin sınıflandırma görevlerinde etkili sonuçlar alabilmek adına, metinler üzerinde çeşitli ön işleme adımları gerçekleştirilmektedir. Metin verisi ön işleme, doğal dil işleme (NLP) ve metin madenciliği görevlerinde verinin kalitesini ve analiz sonuçlarının doğruluğunu artırmak için ham metin verisine uygulanan bir dizi işlemidir. Bu işlemler, metnin küçük harflere dönüştürülmesi, noktalama işaretlerinin kaldırılması, gereksiz kelimelerin (stop words) elenmesi ve kelime köklerinin çıkarılması gibi adımları içermektedir.

Söz konusu doğal afetlerin yaşanması durumunda, hızlı ve etkili sonuçlar elde etmek için elimizde yeterli sayıda veri olsa bile bu verilerin çok az bir kısmı etiketli ve kullanılabilir halde bulunmaktadır. Bu tür durumlarda, elimizde çok fazla veri olmasına rağmen etiketli veri sayısının artırılmasının hem zaman hem de iş gücü açısından oldukça maliyetli olacağı

düşünüldüğünde; etiketli veri sayısının az olduğu durumlarda da başarılı sonuçlar elde etme ihtiyacı doğmuştur. Bu ihtiyaç doğrultusunda Transfer Öğrenme yöntemleri günümüzde yaygın olarak kullanılmaktadır. Başlıca kullanılan yöntemler arasında BERT, RoBERTa ve DistilBERT gibi modeller bulunmaktadır. Bu modeller, geniş veri setlerinde ön eğitim aldıkları için, afetle ilgili sosyal medya metinlerinin gibi spesifik konuların semantik içeriğini anlama ve sınıflandırma konusunda oldukça yeteneklidirler.

Ülkemizin bulunduğu coğrafi konum nedeni ile doğal afetlerin sık yaşanması ve doğal afet sonrasında yaşanan iletişim problemleri göz önüne alındığında; en etkili iletişim kaynağı olan sosyal medyanın daha anlamlı ve etkili kullanılması büyük önem taşımaktadır. Bu doğrultuda yakın zamanda yaşadığımız en büyük depremlerden biri olan 6 Şubat 2023 Türkiye ve Suriye depremine yönelik bir çalışma yapılmıştır.

Çalışma kapsamında az sayıda etiketli veri kullanılarak Transfer Öğrenme yöntemlerinin gücünden yararlanılarak hızlı ve etkili sonuçlar alınmaya çalışılmıştır. Deneysel çalışmalar için daha önceden hazırlanmış olan 6 Şubat 2023 depremine ait Twitter mesajlarının bulunduğu bir veri seti ele alınmıştır (Kaggle deprem veri seti). Bu veri seti içerisinden 2000 tane Türkçe veri seçilmiş ve bu veriler belirlenen sınıflara göre etiketlenmiştir. Sınıflar ilgisiz (0), enkaz (1) ve ihtiyaç (2) olmak üzere 3'e ayrılmıştır. Veri setinde etiket dağılımı ise 654 enkaz, 653 ihtiyaç ve 693 ilgisiz olacak şekilde dengeli olarak belirlenmiştir. Ardından hazırlanan bu etiketli veri seti ile Transfer Öğrenme yöntemlerinden faydalanarak çeşitli deneyler yapılmış ve sonuçlar yorumlanmıştır. Transfer Öğrenme yöntemlerinin başında gelen BERT, DistilBERT ve RoBERTa ile hazırlanan veri setinin tamamı kullanılarak yapılan çalışmaların yanı sıra veri setinin sadece çok az bir kısmı kullanılarak da yine Transfer Öğrenmenin güçlü modellerinden birisi olan few-shot (birkaç örnek ile öğrenme) ve TARS ile de deneyler yapılmış ve sonuçlar kıyaslanmıştır. Tüm bunların yanında geleneksel makine öğrenmesi yöntemlerinden Destek Vektör Makinesi (Support Vector Machine), K-En Yakın Komşu (K-Nearest Neighbors (KNN)), Naive Bayes ve Lojistik Regresyon algoritmalarına da yer verilmiştir. Tüm bu alınan sonuçlar incelendiğinde afet sonrası eldeki oldukça az etiketli veri ile sosyal medyadaki metin verilerinin başarılı bir şekilde sınıflandırılabileceğine dair umut verici sonuçlar gözlemlenmiştir.

#### **Tezin Katkıları:**

- Etiketli bir veri seti elde edilmiştir.
- Bir afet sonrası sosyal medyadaki bilginin hızlı ve doğru bir şekilde sınıflandırılması ve bilginin doğru kullanılması hedeflenmiştir. Bu doğrultuda başarılı sonuçlar alınmıştır.

- Geleneksel yöntemler, Transfer Öğrenme yöntemleri ile bunların yanında TARS ve few-shot (birkaç örnek ile öğrenme) gibi modeller ele alınarak başarımların sonuçları değerlendirilmiştir.
- Benzer çalışmalarda sınıflandırma konusu olarak deprem ele alındığında genel olarak duygu analizi yapılmaktadır. Bu çalışmada farklı bir yaklaşım sağlanarak etiket sayısı 3 olarak seçilmiş ve etiketler arası ayrım ince nüanslara bağlıdır.
- Seçilen modeller ile yapılan deneylerde en yüksek %87,25'lik bir doğruluk oranı elde edilmiştir.
- Bir doğal afet anında sosyal medyanın doğru analiz edildiğinde potansiyel öneminin kanıtlanması için örnek bir çalışma sunulmuştur.



## KURAMSAL TEMELLER

Bu tez çalışmasında, herhangi bir doğal afet (deprem) anında sosyal medyadaki (Twitter) bilginin en hızlı şekilde toplanarak sınıflandırılması ve bu sınıfların incelenmesi konu alınmıştır. Ülkemizin deprem kuşağında yer alması ve aktif fay hatlarına sahip olması nedeniyle ülke olarak depreme her an hazırlıklı olmamız gerekmektedir. Bunun yanı sıra bir deprem sonrası en hızlı şekilde aksiyon alabilmek içinde hazırlıklı olmakta fayda vardır. Bunun en büyük örneklerinden biri olan 6 Şubat 2023 depremi ele alınmıştır. Bu deprem meydana geldiği andan itibaren gerek enkaz altında kalan vatandaşların kendileri veya yakınları tarafından sosyal medya aracılığı ile bildirilmiştir ve yardım ulaştırılmaya çalışılmıştır. Bu yardım sadece enkaz odaklı değil aynı zamanda dışarıda kalan depremzedeler içinde insani ihtiyaçların bildirilmesinde ve ulaştırılmasında büyük önem taşımaktadır.

Sosyal medya, afet durumlarında sadece iletişimi sürdürmekle kalmaz, aynı zamanda afet bölgelerindeki iletişimi de optimize etmektedir. American Red Cross tarafından yapılan bir internet araştırmasına göre, insanlar bilgi alışverişi için sosyal medyayı artan bir şekilde tercih etmektedirler. Amerika'nın %24'ü ve online bireylerin %31'i, ailelerine ve yakınlarına güvende olduklarına dair bilgi vermek için sosyal medya platformlarını kullanmaktadır. (American Red Cross, 2010). Bunun yanında, başka bir araştırma sonucuna göre, afet yaşamış bireyler veya onların yakınları, bulundukları bölgelerin sosyal medya içeriklerinde gösterildiğini gördüklerinde geleceğe dair daha umutlu hissederek daha fazla destekleniyor hissine kapılmaktadırlar (Yazıcı & Zincir, 2014). Bu bağlamda, afetle ilgili bilgi paylaşımında "Twitter" gibi mikroblog platformlarının kullanılması, daha geniş bir kitleye hitap etmek ve bu kitleyle etkin iletişim kurmak için önemlidir.

Twitter (mikroblog platform), sosyal ağların küçük boyutlu bloglar ile birleştirilmiş halidir; küçük ölçekli içerik anında mikrobloglar üzerinden paylaşılabilir ve büyük kitlelere ulaşabilmektedirler (Yazıcı & Zincir, 2014). Fakat burada ele alınması gereken önemli bir konu vardır. Yaşadığımız dönemin getirisi olarak internet hemen hemen herkesi ulaşabildiği bir teknoloji olduğu için dolayısıyla sosyal medya platformları da aynı şekilde oldukça fazla ulaşılabilir haldedir. Bu ulaşılabilirlik aslında bilginin daha hızlı yayılmasını sağlarken aynı zamanda bilgi kirliliğine de yol açmaktadır. Bu nedenle sosyal medya üzerinden bir afet ile ilgili edinilecek bilgilerin sayısı çok fazla olurken netliği de az olacaktır. Bu bilgi kirliliğinin

önüne geçilerek, afet sonrası hızlı bir şekilde verinin toplanması ve sınıflandırılması sonrasında bu sınıflar üzerinden gerekli aksiyonların alınması için bu çalışma yapılmıştır.

Daha önce yapılan çalışmalar incelendiğinde doğal afetler ile ilgili oldukça fazla çalışma yapıldığı görülmüştür. Fakat yapılan çalışmaların bir kısmı geleneksel makine öğrenmesi yöntemleri ile yapıldığı için oldukça fazla sayıda etiketli veriye ihtiyaç duyulmuştur. Bu verilerin toplanıp ardından etiketlenmesinin maliyeti açısından bakıldığında afet sonrası hızlı aksiyon için uygun değildir. Hızlı aksiyon alabilmek için daha az veri ile daha az etiketleme işlemi ile başarılı sonuçlar elde etmek önemlidir. Bu amaç doğrultusunda araştırıldığında transfer öğrenme yöntemlerinin daha uygun olduğu görülmüştür. Transfer öğrenme yöntemleri ile az sayıda etiketli veri ile eğitim yapılarak veya daha önce eğitilmiş makineler üzerinde ince ayar (fine-tuning) yaparak başarılı sınıflandırma sonuçları alındığı gözlemlenmiştir.

Bu tez çalışması, makine öğrenmesinin çeşitli alanlarında kullanılan teknikleri incelemekte ve bu tekniklerin çeşitli uygulamalarına ışık tutmaktadır. Çalışma, geleneksel makine öğrenmesi yöntemleri, transfer öğrenme ve birkaç atış (few-shot) ve sıfır atış (zero-shot) öğrenme tekniklerini kapsamaktadır. Geleneksel makine öğrenmesi yöntemleri iki ana kategoriye ayrılır: denetimli ve denetimsiz öğrenme. Denetimsiz öğrenmede amaç veri集中的 örneklerin çıkışları bilinmediği için tanıma veya sınıflandırma değildir. Genellikle kümeleme, olasılık yoğunluk tahmini, öznitelikler arasındaki ilişkilerin bulunması ve boyut indirgeme gibi amaçlarla kullanılmaktadır (Bilgin, M. (2017)). Bu çalışmada, denetimli öğrenme algoritmalarından Naive Bayes, K-En Yakın Komşu (KNN), Destek Vektör Makinesi (SVM), Lojistik Regresyon ve Karar Ağaçları ele alınmıştır. Transfer öğrenme ise bir modelin öğrendiği bilgileri farklı görevlerde kullanabilmesini sağlar. BERT, DistilBERT ve RoBERTa gibi ön-eğitilmiş dil modelleri, bu çalışmada transfer öğrenme bağlamında kullanılmıştır. Devlin ve arkadaşlarının (2018) BERT üzerine yaptıkları çalışma, bu modelin doğal dil işleme görevlerindeki etkinliğini göstermektedir. Tüm bunlara ek olarak few-shot ve zero-shot öğrenme teknikleri incelenmiştir. Few-shot öğrenme, her sınıf için sadece birkaç etiketli örneği kullanarak yeni sınıfları öğrenme ve sınıflandırma yaklaşımıdır. Zero-shot öğrenme ise, hiç görülmemiş veriler üzerinde çalışabilen modeller geliştirmeyi hedefler. Wang ve arkadaşları (2020), zero-shot öğrenmenin etkinliğini ve uygulama alanlarını detaylı bir şekilde tartışmışlardır. Bu tez çalışmasında kullanılan öğrenme yöntemleri ve algoritmalar, eğitim setlerinin yapısı, model performansları ve uygulama alanları açısından detaylı bir şekilde incelenmiştir. Bu inceleme, her bir modelin avantajlarını ve dezavantajlarını göstermekte ve bu modellerin mevcut probleme nasıl uygulanabileceğine dair bilgiler sunmaktadır.

## Doğal Dil İşleme

Doğal Dil İşleme (NLP) birçok dijital platformdan alınan metin ve ses verilerini otomatik olarak işlemek ve yorumlamak için tasarlanmış bilgisayar bilimi ve dil bilimi arasındaki kesişme noktasında bulunmaktadır. Doğal dil işleme, doğal dilde oluşturulmuş metinler üzerinde otomatik olarak gerçekleştirilen metin ayrıştırma, metin sınıflandırma, bilgi çıkarımı, duygu analizi gibi birçok önemli konuyu bünyesinde barındırmaktadır ve uygulama alanları da gün geçtikçe artmaktadır (Küçük & Arıcı, 2018). Günümüzün dijital çağında, sosyal medya, blog yazıları, forumlar ve diğer online kaynaklardan elde edilen büyük miktarda metin verisi, NLP'nin önemini daha da arttırmıştır. Bu durum, metni derinlemesine anlama, bilgi çıkarımı, duygu tespiti ve otomatik özetleme gibi bir dizi NLP görevinin evrimine katkıda bulunmuştur. Doğal Dil İşleme (NLP) üzerinde çalışırken genellikle iki temel analize odaklanılmaktadır. Bunlar sözdizimsel ve anlamsal analizlerdir.

**Sözdizimsel Analiz:** Bu analiz, bir metni nasıl ayrıştıracağımıza dair dil kurallarını inceleyen süreçtir. Daha basit bir deyişle, bir cümledeki kelimelerin sırasını, kelimelerin cümlede nasıl bir araya geldiğini incelemektedir. Bu, metni anlamlı bir şekilde bölme veya dil bilgisi kurallarına uygun olarak düzenleme çalışmasını içermektedir. Örneğin, "Ali kitap okudu." cümlesinde, "Ali" özne, "kitap" nesne ve "okudu" fiil olarak tanımlanmaktadır.

**Anlamsal Analiz:** Bu, bir metindeki kelimelerin veya cümlelerin ne anlama geldiğini keşfetmeye çalışan süreçtir. Bir kelimenin ya da ifadenin, bağlam içinde ne anlama geldiğini, hangi bilgi veya duyguyu taşıdığını incelemektedir. Örneğin, "yüz" kelimesi, bir insana veya hayvana ait uzuv mu yoksa sayısal bir şeyi mi ifade ediyor? Bu, anlamsal analizin çözmeye çalıştığı türden bir sorundur.

NLP'nin bu iki ana yaklaşımı, makinelerin dilin karmaşıklığını nasıl anladığı ve modellediği üzerine derin bir araştırmadır. İlk olarak, bir veri kümesi (metinler) toplanır. Daha sonra bu veri, yukarıda açıklanan sözdizimsel ve anlamsal analiz süreçlerinden geçirilir. Son aşama, bu metinleri makinelerin anlayabileceği sayısal vektörlere dönüştürmektir. Bu vektörler, daha sonra çeşitli NLP görevlerinde (örneğin, metin sınıflandırma, duygu analizi) kullanılır.

Özetle, NLP'nin amacı, dilin hem yapısal hem de semantik zenginliklerini makineler aracılığıyla modelliyor ve bu bilgileri pratikte kullanmaktır. Teknolojik ilerlemelerle birlikte, bu alandaki yöntemler de sürekli olarak gelişmeye devam etmektedir.

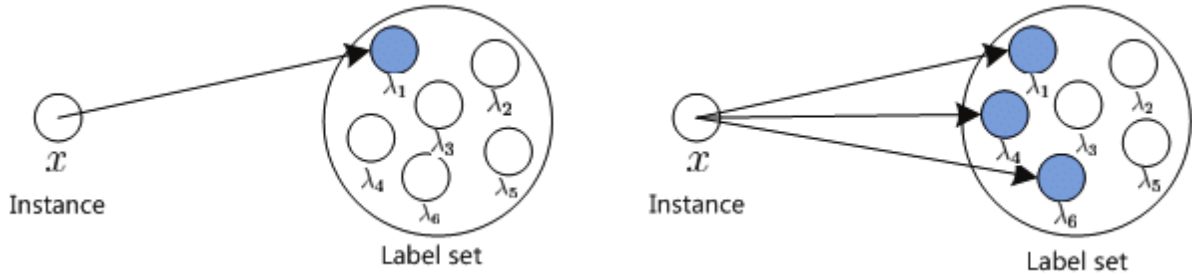
## Metin Sınıflandırma

Metin sınıflandırma, doğal dil işleme (NLP) alanında yer alan, metinleri - cümleler, sorgular, paragraflar ve dokümanlar gibi - belirlenmiş etiketlere atama sürecidir. Bu süreç, spam

tespitinden duygu analizine, haber kategorizasyonundan kullanıcı niyetinin belirlenmesine kadar geniş bir uygulama alanına sahiptir. Metin verileri çeşitli kaynaklardan, örneğin web sitelerinden, e-postalardan, sosyal medyadan veya müşteri hizmetlerinden gelmektedir.

Metin sınıflandırması, manuel etiketleme veya otomatik etiketleme yoluyla yapılmaktadır. Otomatik metin sınıflandırma, endüstrideki metin verilerinin büyüklüğünün artmasıyla giderek daha fazla önem kazanmaktadır (Minaee, Shervin, *et al.* 2021). Otomatik sınıflandırma yöntemleri, kural tabanlı ve makine öğrenimi tabanlı olmak üzere iki ana kategoriye ayrılır. Kural tabanlı yöntemler, belirli kurallar seti kullanarak metni kategorilere ayırırken, makine öğrenimi tabanlı yöntemler, verilerin gözlemlerinden yola çıkarak sınıflandırma yapmayı öğrenmektedir. Makine öğrenimi algoritmaları, önceden etiketlenmiş veri örneklerini kullanarak, metinler ve etiketler arasındaki ilişkileri öğrenmektedir. Bu etiketler tek etiketli veya çok etiketli olmak üzere ikiye ayrılmaktadır.

Tek etiketli sınıflandırma (single-label), var olan  $m$  sınıftan her bir metin için ilgili olduğu belirlenen sadece bir tanesinin seçildiği ve belirlenen sınıfa ait etiketin metne “yapıştırıldığı” ya da “atandığı” uygulamadır. Çok etiketli sınıflandırma (multi-label) ise,  $m$  adet sınıftan bir metin için ilgili olduğu belirlenen bir ya da birden fazla etiket seçilmektedir (Tantuğ, 2016).



**Şekil 1.** Tek etiketli ve çok etiketli metin sınıflandırma örneği (Yangming, *et al.* 2014)

Bu çalışma kapsamında veri ve etiketler düşünüldüğünde ihtiyaç, enkaz, ilgisiz olmak üzere üç kategoride tek etiketli biçimde sınıflandırılacaktır.

Metin sınıflandırması, doğal dil işleme (NLP) alanındaki en temel sorunlardan biridir ve bu süreçte öznitelik (feature) çıkarımı, metin verilerinden anlamlı bilgileri elde etmek için kritik bir adımdır. Öznitelik çıkarımı, ham metin verilerini makine öğrenimi modellerinin anlayabileceği ve işleyebileceği sayısal formata dönüştürmek için kullanılan bir dizi yöntemi kapsamaktadır.

Kelime Torbası (Bag of Words - BoW), metindeki her bir kelimenin frekansını sayarak bir vektör oluşturmaktadır. Bu yöntemde, metnin yapısı ve kelime sıralaması göz ardı edilerek, yalnızca kelime varlıkları ve frekansları dikkate alınmaktadır. Örneğin, "the cat sat on the mat"

cümlesindeki her kelimenin sayısı bir vektörde temsil edilir. Bu yaklaşım, basit ve etkili olmasına rağmen, kelime bağlamını ve sıralamasını göz ardı etmesi nedeniyle bazı sınırlamalara sahiptir.

TF-IDF (Term Frequency-Inverse Document Frequency), bir kelimenin bir metindeki önemini, o kelimenin sıklığına ve tüm belge koleksiyonundaki nadirliğine göre değerlendirmektedir. Bu yöntem, bir kelimenin belirli bir dokümanda ne kadar önemli olduğunu hesaplayarak BoW'un aksine her kelimenin ağırlığını belirlemektedir. Genellikle, bir belgede sık geçen ancak tüm belge kümesinde nadir olan kelimeler yüksek TF-IDF değerlerine sahiptir. Bu, yaygın kelimelerin (örneğin, "ve", "veya" gibi bağlaçlar) aksine, belirli bir belge veya belge kümesi için özel olan terimlerin belirlenmesine yardımcı olmaktadır. TF-IDF, belirli bir belgede önemli anahtar kelimeleri belirlemek ve belge içeriğinin otomatik olarak özetlenmesi gibi işlemlerde sıkça kullanılmaktadır (Ramos, J., 2003).

Kelime Gömme (Word Embedding) yöntemleri, kelimeleri düşük boyutlu sürekli vektörler olarak temsil edilmektedir. Bu vektörler, kelimelerin semantik benzerliklerini yansıtacak şekilde tasarlanmıştır. Word2Vec, GloVe ve FastText gibi modeller, kelimeleri vektör uzayında temsil ederek, kelime anlamlarının ve bağlamlarının daha iyi yakalanmasını sağlamaktadır. Örneğin, Word2Vec modeli, benzer bağlamlarda kullanılan kelimelerin vektör uzayında birbirlerine yakın olmalarını sağlar.

Derin öğrenme tabanlı yaklaşımlar, son yıllarda, derin öğrenme modelleri öznitelik çıkarımında önemli bir yer tutmaktadır. Özellikle, LSTM (Long Short-Term Memory) ve GRU (Gated Recurrent Unit) gibi tekrarlayan sinir ağları (RNN) ve Transformer tabanlı modeller (örneğin BERT, GPT), metin verilerinden öznitelik çıkarmada oldukça başarılıdır. Bu modeller, metnin anlamını daha derinlemesine anlamak için kelime sıralaması ve bağlamı dikkate almaktadır.

## **Veri Seti Toplama**

Veri seti toplama, belirli bir amacı veya araştırmayı desteklemek üzere çeşitli kaynaklardan bilgi toplama sürecidir. Bu, modern bilimde ve endüstride önemli bir adımdır; çünkü doğru ve yeterli veriye sahip olmak, analizlerin, modellemenin ve tahminlerin kalitesini büyük ölçüde etkilemektedir.

Veri setlerinin toplanmasının temel nedeni, bilgiye dayalı kararlar almak ve bilimsel araştırmalar için gerçekçi temeller oluşturmaktır. Örneğin, bir sağlık kuruluşu, ilaçların etkinliğini değerlendirmek için klinik deney verilerini toplayabilirken, bir perakende şirketi ise müşteri alışkanlıklarını ve tercihlerini anlamak için satış verilerini inceleyebilmektedir.

Toplanacak verinin ne için kullanılacağını belirlemek esastır. Bu, hangi tür verinin toplanacağını ve hangi yöntemlerin kullanılacağını belirlemektedir. Bu çalışma kapsamında sosyal medya ve deprem yönetimi arasındaki ilişkiyi anlamak için veri setinin içeriği Twitter mesajlarından oluşmaktadır. 6 Şubat 2023 depreminin olduğu andan itibaren Twitter API aracılığı ile toplanan 478052 tweet 'in olduğu bir set hazır olarak bulunmaktadır (Kaggle deprem veri seti). Bu set üzerinden çalışmanın ihtiyaçları doğrultusunda yeni küçük veri setleri oluşturulmuştur.

## Veri Etiketleme

Veri etiketleme, genellikle makine öğrenimi modelleri için eğitim verisi oluşturmak amacıyla, ham verilere anlamlı etiketlerin atanması işlemidir. Bu etiketler, metin, görüntü, ses veya video gibi farklı veri türlerinden oluşmaktadır.

Veri etiketleme işlemi yapılırken dikkat edilmesi gereken bazı önemli noktalar vardır. Bunlar:

- Doğruluk: Etiketleme işleminin en önemli unsuru doğruluktur. Yanıltıcı etiketler, modelin yanlış öğrenmesine ve yanlış sonuçlar üretmesine neden olmaktadır.
- Tutarlılık: Etiketleme sürecinde, veriler arasında tutarlılık olmalıdır. Aynı türdeki verilere aynı etiketin atanması gerekmektedir.
- Ayrıntıcılık: Özellikle metin sınıflandırma gibi konularda, veriyi etiketlerken yeterince ayrıntılı olmak önemlidir. Örneğin, bir tweetteki duygu "olumlu", "olumsuz" ya da "nötr" olarak basitçe etiketlenebilirken, daha ayrıntılı etiketler (örn. "çok olumlu", "hafif olumlu" vs.) modelin daha spesifik sonuçlar üretmesine olanak tanımaktadır.
- Öznellik Sorunu: Özellikle sosyal medya yorumlarını etiketlerken, insanların subjektif değerlendirmeleri etiketleme sürecini zorlaştırmaktadır. Bu nedenle, birden fazla kişi tarafından etiketlenen verilerin karşılaştırılması ve fikir birliği oluşturulması önemlidir.

Çalışma kapsamında bu belirtilen önemli noktalara dikkat edilerek bir etiketleme işlemi yapılmıştır. Tüm bunların yanında çalışma kapsamında kullanılan veri seti 3 farklı değerlendirici tarafından etiketlenmiş olup, ardından bu 3 farklı değerlendirme için Fleiss' Kappa değeri hesaplanmıştır.

## Veri Ön İşleme

Metin tabanlı bir belge alfabetik karakterler haricinde emoji, rakam, link, kullanıcı adı, kısaltmalar, edat/bağlaç, noktalama işaretleri vb. içeriklerden oluşmaktadır. Bu içerikler metin sınıflandırma amacına bağlı olarak bir anlam ifade etmeyebilir (Salur & Aydın, 2018). Bu nedenle, metni makine öğrenimi modelleri için uygun bir forma getirme süreci olan veri ön işleme aşaması kritik bir öneme sahiptir.

Veri ön işleme, ham metin verisinin analize uygun bir forma dönüştürülmesi sürecidir. Bu süreç, başlangıçta metinde yer alan ama analiz için gereksiz olan bilgilerin temizlenmesini içerir. Örneğin; URL'ler, kullanıcı adları ve gereksiz noktalama işaretleri gibi ögeler metin analizi için genellikle yararsızdır ve bu nedenle kaldırılarak daha anlamlı hale getirilmektedir. Bunun yanında, tutarlılık ve standardizasyon sağlamak için metindeki tüm harfler küçük harfe dönüştürülmektedir. Ardından, metin, tokenizasyon süreciyle kelimelere veya tokenlara ayrılır, bu sayede metin üzerinde daha detaylı işlemler yapılabilir hale gelmektedir.

Bir sonraki adım, genelde metin içerisinde sıkça bulunan ancak spesifik bir anlam taşımayan stop kelimelerinin kaldırılmasıdır. Bu kelimeler, 've', 'ama', 'ile' gibi yaygın Türkçe bağlaçlar olarak örneklendirilebilir. Daha sonra, kelimeler lemmatizasyon veya kök bulma süreçlerinden geçirilerek temel köklerine indirgenir. Özellikle Twitter gibi platformlarda sıkça karşılaşılan kısaltmaların ve jargonların tam halleriyle değiştirilmesi de önem taşımaktadır. Belirtilen bu adımlar, metin sınıflandırma projelerinde genellikle uygulanan prosedürlerin bir özetidir. Ancak, her projede bu adımların özel ihtiyaçlara göre uyarlanması gerekmektedir. Sonuç olarak veri ön işleme adımları yapılan çalışmanın amacına, çalışmadaki verinin içeriğine göre değişiklik göstermektedir.

Bu çalışma kapsamında deprem ile ilgili Twitter mesajları düşünüldüğünde, belirli terimlerin, hashtag'lerin veya emoji kullanımlarının da dikkate alınması da önemlidir. Tablo 1'de bu çalışmaya göre belirlenmiş ön işleme adımlarının bir tweet üzerinde uygulanmasına dair örnek verilmiştir.

**Tablo 1.** Örnek Bir Tweet Üzerinde Ön İşleme

| Yapılan İşlem                                                                                    | Tweet                                                                                                                                                                                                                                                                                                                                       |
|--------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Tweet'in ilk hali</b>                                                                         | YARDIM!!! Ballica mahallesi. 11.002 sokak Hicret apartmanı /Kahramanmaraş Hatice Yılmaz Göçük altında hala AFAD ekipleri gitmemiş bulunmaktadır acil yardımınız lazim vücudu buz tutmuş arada ses veriyormuş 😞 @haluklevent @OguzhanUgur #deprem #Turkey #Kahramanmarasdeprem <a href="https://t.co/8FyihXIJ9e">https://t.co/8FyihXIJ9e</a> |
| <b>Hashtag ve @ ile Başlayan Etiketlerin Kaldırılması</b>                                        | YARDIM!!! Ballica mahallesi. 11.002 sokak Hicret apartmanı /Kahramanmaraş Hatice Yılmaz Göçük altında hala AFAD ekipleri gitmemiş bulunmaktadır acil yardımınız lazim vücudu buz tutmuş arada ses veriyormuş 😞 <a href="https://t.co/8FyihXIJ9e">https://t.co/8FyihXIJ9e</a>                                                                |
| <b>İnternet bağlantılarının kaldırılması</b>                                                     | YARDIM!!! Ballica mahallesi. 11.002 sokak Hicret apartmanı /Kahramanmaraş Hatice Yılmaz Göçük altında hala AFAD ekipleri gitmemiş bulunmaktadır acil yardımınız lazim vücudu buz tutmuş arada ses veriyormuş 😞                                                                                                                              |
| <b>Emoji ve noktalama işaretlerinin kaldırılması (nokta ve türkçe karakterler korunmalıdır.)</b> | YARDIM Ballica mahallesi. 11.002 sokak Hicret apartmanı Kahramanmaraş Hatice Yılmaz Göçük altında hala AFAD ekipleri gitmemiş bulunmaktadır acil yardımınız lazim vücudu buz tutmuş arada ses veriyormuş                                                                                                                                    |
| <b>Küçük harfe dönüştürme</b>                                                                    | yardım ballica mahallesi. 11.002 sokak hicret apartmanı kahramanmaraş hatice yılmaz göçük altında hala afad ekipleri gitmemiş bulunmaktadır acil yardımınız lazim vücudu buz tutmuş arada ses veriyormuş                                                                                                                                    |

## Geleneksel Makine Öğrenmesi

Geleneksel makine öğrenmesi, genellikle belirli bir görev veya uygulama için özel olarak toplanmış verilere dayanarak model oluşturmayı amaçlar. Örneğin, bir kredi skoru tahmini yapmak için alınan verilere dayalı bir model eğitildiğinde, bu model başka bir finansal görevde doğrudan kullanılamaz. Geleneksel yaklaşımlar, genellikle öznel mükemmellik gerektirir ve modelin eğitildiği veri setinin özelliklerine oldukça bağımlıdır.

Geleneksel makine öğrenmesi yaklaşımları, genellikle her bir problemi bireysel bir bulmaca olarak ele almaktadır. Bu yaklaşımın temelinde, her yeni görevin kendine özgü bir zorlukla geldiği ve bu nedenle sıfırdan, her defasında yeni bir çözüm yolu izlenmesi gerektiği fikri yatmaktadır. Bu, birçok uygulamada etkili sonuçlar vermiştir, ancak makine öğrenmesinin evrimi boyunca bazı sınırlılıklarını da ortaya çıkarmaktadır.

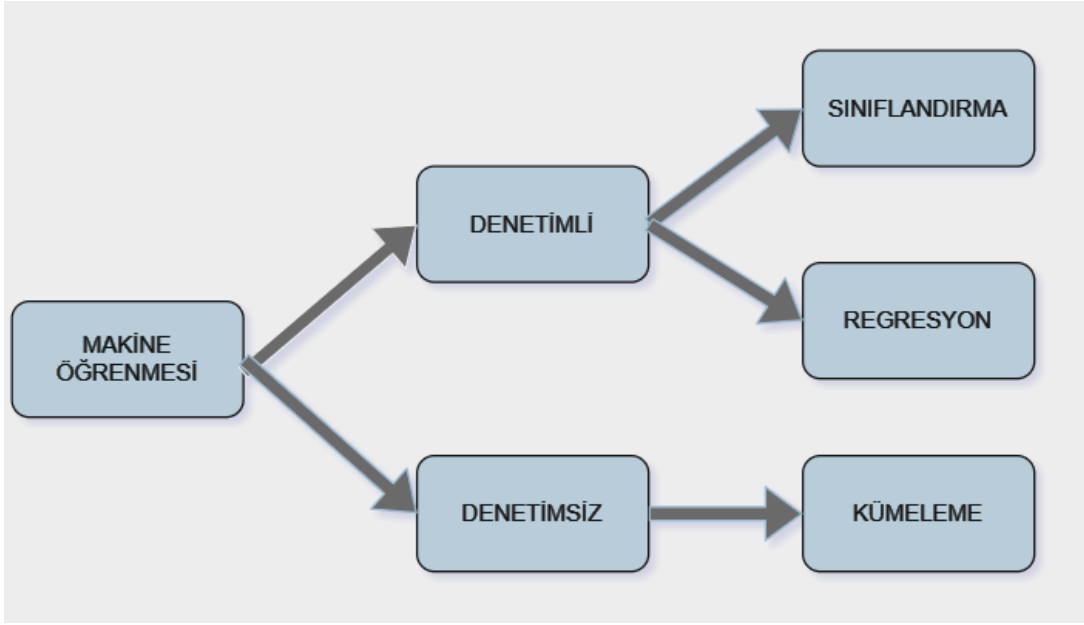
İnsan öğrenmesi önceden karşılaşmadığı problemlere çözüm ararken gerek farkında olarak gerekse farkında olmadan geçmiş deneyimlerinden faydalanmakta ve bu sayede benzerlikler içeren çözümlerden daha etkin çözümler üretmektedir (Barış Koçer, 2012). Örneğin, bir çocuk bir kez sıcak bir nesneye dokunduğunda, bu deneyimi hafızasına kaydeder ve ilerleyen zamanlarda sıcak nesnelere dokunmaktan kaçınır. Bu, insan öğrenmesinin adaptif ve kümülatif bir doğası olduğunu gösterir. Geleneksel makine öğrenmesi bu tür bir kümülatif

öğrenme yaklaşımını takip etmez. Her yeni problem için, geçmiş deneyimlerden bağımsız olarak sıfırdan bir çözüm arayışına girmektedir.

Bu, özellikle veri dağılımının sürekli olarak değiştiği, dinamik ortamlarda problem teşkil etmektedir. Çünkü geleneksel makine öğrenmesi algoritmaları, eğitildikleri veri dağılımı üzerine optimize edilir ve eğer bu dağılım değişirse performanslarında düşüşler gözlemlenmiş olur. Gerçekte, birçok uygulama senaryosunda, veri dağılımları sürekli olarak değişiklik gösterebilir ve bu da geleneksel yöntemlerin bazı durumlarda sınırlı kalmasına neden olmaktadır.

Sonuç olarak, geleneksel makine öğrenmesi yöntemleri belirli görevlerde başarılı sonuçlar alabilirken, değişken ve dinamik ortamlarda veya geçmiş deneyimlerin faydalı olabileceği senaryolarda sınırlılıklarını göstermektedir. Bu sınırlılıkların üstesinden gelmek ve daha esnek çözümler sunmak için makine öğrenmesinde yeni yaklaşımlara ihtiyaç duyulmaktadır.

Geleneksel makine öğrenmesi yöntemlerinin iki ana dalı, denetimli ve denetimsiz öğrenme olarak bilinir ve her biri veriden öğrenme sürecine farklı yaklaşımlar getirmektedir. Denetimli öğrenmede, etiketlenmiş veri setlerinden yararlanılır, bu durumda her veri noktasının hem girdi (örneğin, bir görüntü veya metin) hem de çıktı etiketi (örneğin, bir kategori veya sayısal değer) bulunmaktadır. Bu yaklaşımda, modeller girdi ve çıktılar arasındaki ilişkileri öğrenmek için eğitilir, böylece sınıflandırma veya regresyon gibi görevlerde kullanılabilirler. Karar ağaçları, destek vektör makineleri, rastgele ormanlar ve lojistik regresyon gibi algoritmalar bu kategoride yaygın olarak tercih edilmektedir. Diğer taraftan, denetimsiz öğrenme, etiketlenmemiş veri setleri üzerinde gerçekleşir ve algoritma, verilerin içindeki yapıları, desenleri ve ilişkileri kendi başına keşfetmeye çalışmaktadır. Bu süreçte, kümeleme ve boyut azaltma gibi yöntemler kullanılır, örneğin K-means kümeleme, hiyerarşik kümeleme, DBSCAN ve ana bileşen analizi (PCA). Sağlık, finans, perakende ve telekomünikasyon gibi çeşitli alanlarda, bu makine öğrenimi yöntemleri, veriye dayalı karar verme ve tahmin süreçlerinin daha verimli ve etkili olmasını sağlamaktadır.



**Şekil 2.** Denetimli ve denetimsiz metin sınıflandırma

Bu çalışmada, modelin eğitimi için KNN, Destek Vektör Makinesi (SVM), Naive Bayes, Karar Ağacı ve Lojistik Regresyon gibi sınıflandırma teknikleri tercih edilmiştir. Bu algoritmalar, metinlerin doğru bir şekilde kategorize edilmesi için kullanılmış ve modelin performansını transfer öğrenme yöntemleri ile kıyaslamak amacıyla uygulanmıştır.

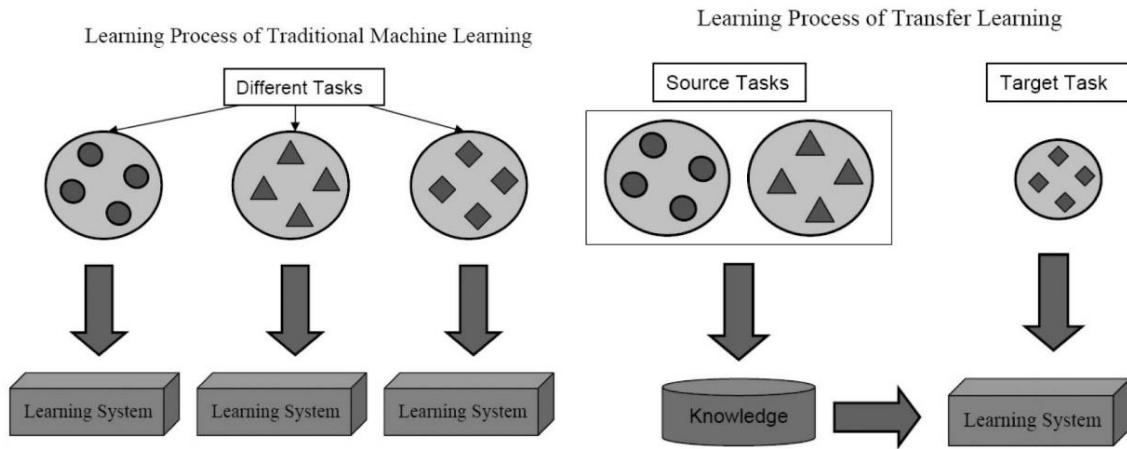
KNN algoritması, bir veri noktasını, en yakın komşularının çoğunluk oyuyla sınıflandırmaktadır. "K", seçilen en yakın komşu sayısını temsil etmektedir. Yeni bir veri noktası geldiğinde, algoritma en yakın K komşuyu bularak ve bu komşuların en yaygın sınıfını tahmin olarak sunmaktadır. K (Komşu Sayısı) ne kadar büyük bir K değeri seçilirse, o kadar geniş bir komşuluk dikkate alınır, bu da modelin genelleme yeteneğini artırabilir veya azaltabilir. SVM, farklı sınıflara ait veri noktalarını en iyi ayıran bir hiper-düzlem oluşturmaktadır. Bu hiper-düzlem, sınıflar arasındaki marjı maksimize ederek optimum bir ayırım sağlamaktadır. Naive Bayes, her özelliğin bağımsız olduğu varsayımı üzerine kuruludur ve Bayes Teoremi'ni kullanarak sınıflandırma yapmaktadır. Verilen özelliklerin sınıf olasılıklarını hesaplayarak, en yüksek olasılığa sahip sınıfı tahmin olarak sunar. Karar Ağacı, veri setini basit karar düğümleri kullanarak bölümlere ayırmaktadır. Her düğüm, bir özelliği kontrol eder ve veri setini daha küçük alt kümelerine bölmektedir. Bu süreç, belirli bir durdurma kriterine ulaşana kadar devam etmektedir. Lojistik Regresyon, bağımsız değişkenler ve iki kategorili bir bağımlı değişken arasındaki ilişkiyi modellemek için kullanılmaktadır. Bu algoritma, bir olasılık skoru vererek sınıflandırma yapmaktadır.

## Transfer Öğrenme Yöntemleri

Transfer öğrenme, makine öğrenmesinin bir alt dalıdır ve temelde öğrenilen bilgilerin bir görevden diğerine aktarılması üzerine kuruludur. Bu yaklaşım, öğrenme sürecinin daha verimli ve hızlı hale gelmesi için önceki deneyimlerden faydalanma ilkesine dayanır. Transfer öğrenmenin arkasındaki fikir, öğrenilen bilginin, bir görev için elde edildiği zaman, diğer benzer görevlerde de kullanılabileceğidir. Bu, insanların yeni beceriler kazanırken geçmişte öğrendikleri bilgileri nasıl kullandığına benzetilmektedir.

Geleneksel makine öğreniminde, her bir problem genellikle izole edilmiş bir şekilde ele alınır ve bu problemler arasında bilgi aktarımı yapılmamaktadır. Ancak, transfer öğrenmede, bir model başka bir görev için eğitildiğinde, bu modelin bilgisi başka bir göreve aktarılabilir. Bu, özellikle az miktarda etiketli veriye sahip olan yeni görevler için büyük bir avantajdır. Modelin, zengin veriye sahip olan başka bir görevde önceden eğitilmesi ve bu öğrenilen bilgilerin yeni görevde uygulanması, modelin performansını büyük ölçüde artırmaktadır.

Ancak, transfer öğrenmenin de kendi zorlukları vardır. Özellikle, farklı görevler arasında yeterince benzerlik olmazsa, yanıltıcı sonuçlara yol açmaktadır. Aktarılan bilginin, yeni görevle ne kadar uyumlu olduğunu belirlemek kritiktir. Ayrıca, hangi bilgilerin aktarılması gerektiğini ve hangi bilgilerin göz ardı edilmesi gerektiğini belirlemek için dikkatli bir değerlendirme yapılmalıdır. Sonuç olarak, transfer öğrenme, makine öğreniminin geleneksel yaklaşımlarına kıyasla birçok avantaja sahip olmasına rağmen, bu yöntemin başarılı bir şekilde uygulanabilmesi için özel dikkat ve değerlendirmenin gerekli olduğu unutulmamalıdır. Bu yaklaşım, öğrenme sürecini hızlandırmada ve daha az veriyle daha etkili sonuçlar elde etmede devrimci bir adımdır.



**Şekil 3.** Geleneksel makine öğrenmesi ve transfer öğrenme (Pan & Yang, 2010)

Ön eğitilmiş dil modelleri doğal dil işleme de görev performansını artırmak için kullanılmaktadır (Houlsby *et al.* 2019). Transfer öğrenmenin popülerliğinin artmasıyla, bu

alandaki birçok alt algoritma ve model geliştirilmiştir. Özellikle doğal dil işleme (NLP) alanında, BERT (Bidirectional Encoder Representations from Transformers) gibi modeller ortaya çıkmıştır. BERT'in başarısının ardından, onun daha hafif ve daha hızlı bir sürümü olan DistilBERT gibi türevleri geliştirilmiştir. Benzer şekilde, RoBERTa da BERT'in bir başka varyasyonu olarak ortaya çıkmıştır ve daha geniş veri setleri üzerinde ve daha uzun eğitim süreleri ile optimize edilmiştir. Bu alt algoritmalar ve modeller, transfer öğrenmenin gücünü artıran ve makine öğrenimi topluluğu tarafından geniş bir şekilde kabul gören yeniliklerden bazılarıdır. Bu çalışma kapsamında ön-eğitilmiş transfer dil modellerinden BERT, DistilBERT ve RoBERTa modellerinin yanı sıra sıfır atış öğrenme (zero-shot), birkaç örnek öğrenme (few-shot) ve TARS gibi modelleri de kullanılarak eğitimler yapılmış ve başarımları sonuçları değerlendirilmiştir.

### **BERT modeli**

BERT (Bidirectional Encoder Representations from Transformers) modeli, Google AI tarafından 2018'de tanıtılan bir doğal dil işleme modelidir. Birçok NLP görevinde o ana kadar elde edilen en iyi performansı aşarak büyük bir etki yaratmıştır. BERT'in başarısının arkasındaki ana nedenlerden biri, metni hem sağdan hem de soldan dikkate alarak çift yönlü bir yaklaşımla metni anlamasıdır. Geleneksel dil modelleri ya metni soldan sağa ya da sağdan sola değerlendirirken, BERT her iki yönde de metni analiz etmektedir. Bu, kelimenin hem önceki hem de sonraki kelimelerle olan ilişkisinin daha iyi yakalanmasını sağlamaktadır. Böylece bağlam daha doğru bir şekilde belirlenmektedir. Önceden eğitim aşamasında, belirli kelimeler metindeki yerlerinden kaldırılır (maskelenir) ve model bu eksik kelimeleri tahmin etmeye çalışarak öğrenir.

BERT, transformer mimarisini kullanır. Transformer, "Attention is All You Need" başlıklı 2017 tarihli makalede Vaswani ve arkadaşları tarafından tanıtılmıştır. Temelde, sıralı veriler üzerinde çalışan yöntemlerin aksine, paralel hesaplamaların avantajlarından yararlanarak veriyi işler.

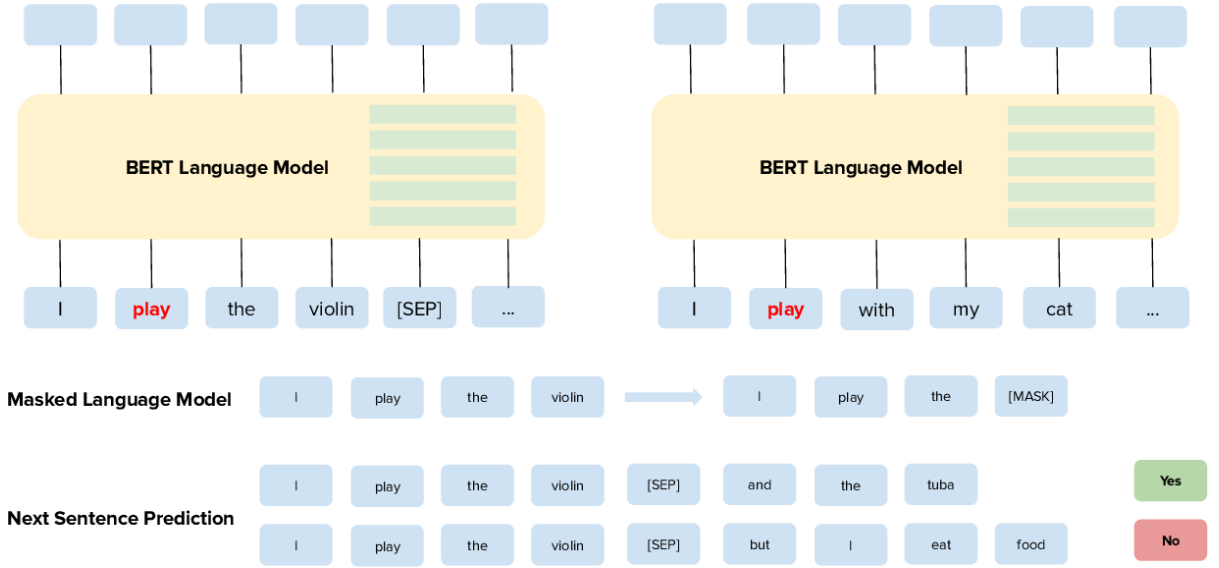
BERT modeli, milyonlarca kelimenin bulunduğu büyük metin yığınları üzerinde eğitilmektedir (Devlin *et al.* 2018). Bu sayede geniş bir kelime dağarcığına ve dil bilgisine sahip olmaktadır. Model, farklı boyutlarda ve farklı yapılandırmalarda mevcuttur. Örneğin, BERT-Base ve BERT-Large olarak iki ana versiyonu bulunmaktadır. Her iki model de Transformer mimarisine dayanmaktadır, ancak kapasite ve eğitim detayları açısından farklılıklar göstermektedirler (Alammar, 2021).

BERT-Base, transformer mimarisinin 12 katmanlı bir sürümüdür. Bu model, görece daha kompakt bir yapıya sahip olup, geniş bir uygulama alanı için dengeli bir seçenek sunmaktadır. Toplamda yaklaşık 110 milyon parametre içerir, bu da onun hızlı eğitim ve tahmin süreçleri sunmasını sağlamaktadır. Ancak, bu kompakt yapı bazen daha karmaşık dil yapıları veya daha zor NLP görevleri için yeterli olmamaktadır. Buna karşın, BERT-Large, adından da anlaşılacağı gibi, çok daha büyük bir modeldir. 24 Transformer katmanı ile iki kat daha fazla derinliğe sahiptir ve yaklaşık 340 milyon parametre içermektedir. Bu ek kapasite, BERT-Large'ın daha karmaşık dil yapılarını ve görevlerini modellemesine olanak tanımaktadır. Bununla birlikte, bu büyüklük, eğitim süresini ve kaynak tüketimini önemli ölçüde artırmaktadır.

BERT, çok sayıda dikkat katmanından (attention layer) oluşan transformer mimarisi kullanır. Her dikkat katmanı, metindeki her kelimenin diğer tüm kelimelerle olan ilişkilerini değerlendirerek bir ağırlık matrisi oluşturmaktadır. Bu ağırlık matrisi, modelin bir kelimenin diğer kelimelerle olan bağlamını anlamasına yardımcı olur. Genellikle büyük miktarda metin verisiyle önceden eğitilir. Özellikle Wikipedia ve BooksCorpus gibi büyük metin kaynakları kullanılarak eğitim gerçekleştirilir. Bu önceden eğitim, modelin genel dil bilgisi ve kelime dağarcığını kazanmasına olanak tanır. Önceden eğitim aşamasında BERT, iki ana görevi temel alarak öğrenme gerçekleştirir. Bunlar MLM (Masked Language Modeling) ve NSP (Next Sentence Prediction) şeklindedir.

MLM, rastgele seçilen kelimeler metinden maskeleme işlemine tabi tutulur. Model, maskelenen bu kelimeleri doğru bir şekilde tahmin etmeye çalışmaktadır. Örneğin, "Kediler fareleri sever." cümlesinde "fareleri" kelimesi maskelendiğinde, BERT bu kelimeyi tahmin etmek için "Kediler [MASK] sever." cümlesini kullanır.

NSP, iki cümle modele verilir ve ikinci cümlenin, birincisinin devamı olup olmadığını tahmin etmeye çalışır. Bu, daha büyük metin parçalarının bağlamını anlama yeteneğini artırır.



Şekil 4. Maskelenmiş dil modeli ve sonraki cümle tahmini (Debora *et al.* 2020)

### RoBERTa modeli

Roberta, Facebook AI tarafından BERT modelinin bir varyasyonu olarak geliştirilmiştir. Roberta'nın adı, "Robustly Optimized BERT Pretraining Approach"ın kısaltmasıdır. BERT ile benzer bir yapıya sahip olmasına rağmen, Roberta'nın eğitimi ve bazı teknik özellikleri BERT'ten farklıdır. Roberta'nın çalışma prensibi; BERT ile benzer bir biçimde transformer mimarisini temel almaktadır. Maskelenmiş dil modeli görevini kullanarak önceden eğitim görür, ancak bu süreçte bazı farklılıklar bulunmaktadır. Özellikle, Roberta'nın eğitimi daha büyük veri kümeleri ve daha uzun süre üzerinde gerçekleştirilir. Bunun yanında, Roberta'nın önceden eğitim görevi sırasında "next sentence prediction" (bir sonraki cümle tahmini) görevi kullanılmaz. BERT'e kıyasla eğitimde dört kat daha fazla dizi görmek için sekiz kat daha büyük bir parti boyutuyla yarısı kadar optimizasyon adımı için eğitilmiştir (Liu *et al.* 2019).

BERT ile Roberta arasındaki başlıca farklar:

- Eğitim Verisi ve Süresi: Roberta, BERT'e kıyasla daha büyük veri kümeleri üzerinde eğitilmiştir. Bu, modelin daha geniş bir dil bilgisine sahip olmasını sağlar.
- Önceden Eğitim Görevleri: BERT, maskelenmiş dil modeli ve bir sonraki cümle tahmini görevleriyle eğilirken, Roberta yalnızca maskelenmiş dil modeli göreviyle eğilir.
- Hiperparametre Optimizasyonu: Roberta'nın geliştirilmesinde, BERT'ten farklı olarak, modelin daha etkili bir şekilde önceden eğitilmesi için hiperparametre optimizasyonu büyük bir rol oynamaktadır.

Sonuç olarak, Roberta, BERT'in başarısını temel alarak bazı değişikliklerle ve optimizasyonlarla daha da ileri taşımayı amaçlayan bir modeldir.

### **DistilBERT modeli**

DistilBERT, BERT'in daha hafif ve hızlı bir sürümüdür. BERT gibi, dil modelleme görevlerinde yüksek performans gösterir, ancak daha küçük ve daha hızlıdır. Bu, BERT'in "damıtılmış" bir versiyonudur, yani orijinal BERT modelinden bilgi aktarımı yapılarak öğrenci bir modele (DistilBERT) eğitilmiştir (Sanh, 2019).

DistilBERT, öğretmen-öğrenci paradigmasını kullanarak eğitilmektedir. Burada, öğretmen model (BERT) zaten eğitilmiştir. Öğrenci model (DistilBERT), öğretmen modelinden bilgi alarak eğitilir. DistilBERT, öğretmen modelin çıktılarına olabildiğince yakın çıktılar üretmeye çalışmaktadır. Bu damıtma işlemi sonucunda, DistilBERT modeli BERT'in orijinal boyutunun yarısı kadar olmasına rağmen, görev performansında sadece küçük bir düşüş göstermektedir.

DistilBERT, BERT'in orijinal boyutunun yaklaşık yarısı kadardır. BERT'e göre daha hızlıdır, çünkü daha az parametreye sahiptir. Çeşitli görevlerde BERT'in performansına oldukça yakın sonuçlar vermektedir. Ancak mutlak performansta genellikle bir miktar düşüş vardır. DistilBERT, öğretmen-öğrenci paradigmasını kullanarak BERT'ten bilgi aktarımı ile eğitilirken, BERT ve RoBERTa kendi başlarına eğitilmektedirler. DistilBERT, BERT'in avantajlarını koruyarak daha hafif ve hızlı bir modeldir. Uygulamalar için modelin boyutu ve hızı önemliyse (örneğin mobil uygulamalar veya tarayıcı tabanlı uygulamalar), DistilBERT ideal bir seçenek olmaktadır.

### **Zero-Shot (Sıfır Atış) öğrenme yöntemi**

Sıfır Atış Öğrenme (Zero-Shot Learning), makine öğrenimi alanında, modelin eğitim sırasında hiç görmemiş olduğu sınıfları tanıyabilme yeteneğini ifade etmektedir. Bu yöntem, özellikle veri kısıtlı durumlarda veya sürekli değişen veri kümelerinde büyük önem taşımaktadır. Sıfır Atış Öğrenme, modelin mevcut sınıflar arasındaki ilişkileri genelleştirmesini ve bu bilgiyi daha önce hiç karşılaşmadığı yeni sınıflara uygulamasını gerektirir. Sınıfların veya görevlerin tanımlarının mevcut olduğu ve etiketlenmemiş veri eksikliğini telafi etmek için kullanılacağı varsayılmaktadır. Bu durum, başlangıçta az eğitim verisi olması nedeniyle basitçe ortaya çıkabileceği gibi, sınıfların veya görevlerin sayısının çok fazla olması nedeniyle de meydana gelebilmektedir (Larochelle *et al.* 2008).

Bu alandaki temel yaklaşımlar genellikle, sınıflar arası paylaşılan ve transfer edilebilen semantik özelliklere dayanmaktadır. Örneğin, bir model "kanatlı" veya "dört ayaklı" gibi genel

özellikleri kullanarak tanımadığı hayvan türlerini tanıyabilir. Bu yaklaşım, görsel tanıma, metin sınıflandırma ve konuşma tanıma gibi çeşitli alanlarda uygulanmaktadır.

Sıfır Atış Öğrenme, makine öğrenimi modellerinin esnekliğini ve uygulanabilirliğini artırma potansiyeline sahip bir alan olarak kabul edilmektedir. Sürekli genişleyen ve değişen veri kümelerine adaptasyon sağlamada önemli bir rol oynayan bu yaklaşım, modellerin gerçek dünya senaryolarında daha etkin performans göstermesine olanak tanımaktadır.

### **Few-Shot (Birkaç Atış) öğrenme yöntemi**

Few-shot ya da birkaç örnek ile öğrenme, yapay zekâ ve makine öğrenimi alanlarında kullanılan bir yöntemdir. Bu yaklaşım, modelin çok az sayıda örnek üzerinden öğrenmesini sağlamaktadır. Bu genellikle yeni görevler veya kavramlar için gerekmektedir. Çünkü her zaman yeterli miktarda eğitim verisi bulmak mümkün olmazken hem de yeterli süre sorunu da oluşmaktadır. İnsanların bilgi birikimleri sayesinde az sayıda örnekle yeni şeyler kolayca öğrenebilmesinden esinlenerek az sayıda örnekle öğrenme yöntemi önerilmiştir (Sui *et al.* 2020). Few-shot öğrenmenin temel amacı, modelin genelleme yeteneğini artırarak, sınırlı veri ile bile etkili bir şekilde öğrenebilmesini sağlamaktır.

Bu tür öğrenme süreci genellikle önceden büyük veri setleriyle eğitilmiş modellerin temel alınarak gerçekleştirilmektedir. Modelin önceden kazandığı bilgi ve deneyimler, yeni ve daha az verili durumlara uygulanmaktadır. Transfer öğrenme, bu sürecin temel bir parçasıdır ve modelin önceden öğrendiği özellikleri yeni görevlere uyarlamasını içermektedir. Few-shot öğrenme metodolojisi, modelin önceden öğrenilmiş genel bilgilerini ve bu bilgileri yeni, spesifik görevlere nasıl uygulayacağını öğrenmesini kapsamaktadır. Bu, modelin yeni veriler üzerinde hızlı bir şekilde öğrenmesini ve uyum sağlamasını sağlamaktadır.

Her bir örnek few-shot öğrenme için çok önemlidir. Model, bu örneklerden maksimum bilgiyi çıkarmaya çalışır, bu da özellikle veri kısıtlı olduğunda önem taşımaktadır. Bu yaklaşımın zorlukları arasında, modelin aşırı uydurma yapma riski ve sınırlı veri nedeniyle genelleme yaparken yaşadığı zorluklar bulunmaktadır. Ayrıca, modelin yeni durumlara nasıl uyum sağlayacağını öğrenmesi de zorluklar arasındadır.

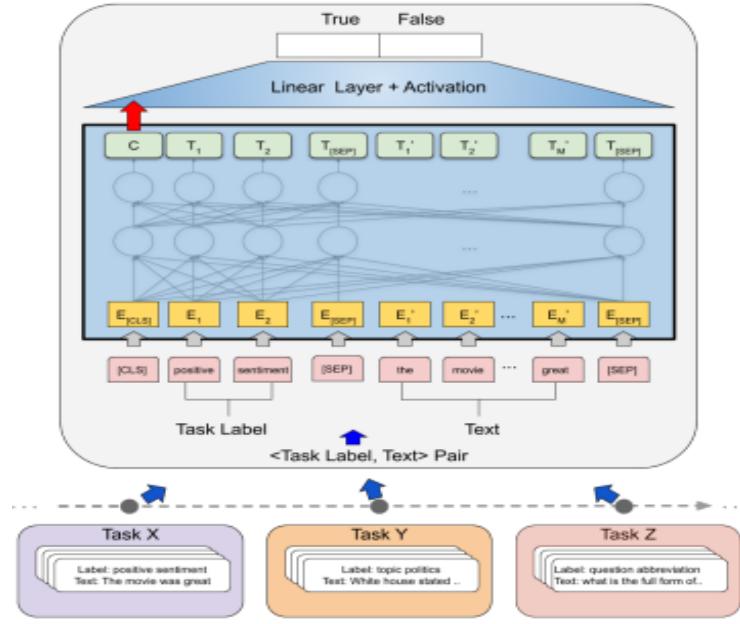
### **TARS modeli**

Metin sınıflandırması, verilen bir metin için bir veya birden fazla sınıf etiketini tahmin etme işlemidir ve spam filtreleme, duygu analizi, niyet tespiti ve haber konusu sınıflandırması gibi birçok alanda kullanılmaktadır. Modern yaklaşımlar, BERT gibi dönüştürücü (transformer) mimarilerini kullanarak bu görevi yerine getirmektedir. Bununla birlikte, az örnekli (few-shot) ve sıfır örnekli (zero-shot) senaryolarda, geleneksel transfer öğrenme yöntemlerinin

sınırlılıkları vardır. TARS (Task-Aware Representation of Sentences), bu sınırlılıkları aşmak için önerilen bir yöntemdir. TARS, herhangi bir metin sınıflandırma görevini genel bir ikili sınıflandırma problemine dönüştürmektedir. Bu yaklaşımda, metin ve sınıf etiketi birlikte değerlendirilmektedir. Böylece, model hem göreve duyarlıdır hem de sınıf etiketinin anlamsal bilgilerini kullanabilmektedir. Bu, özellikle az örnekli ve sıfır örnekli öğrenme senaryolarında avantaj sağlamaktadır (Halder, Kishaloy, *et al.* 2020).

TARS, çapraz dikkat mekanizması kullanarak metin ve etiket arasındaki ilişkiyi öğrenmektedir. Eğitim sırasında, her sınıf etiketi için doğru veya yanlış etiketli örnekler kullanılmaktadır. Bu yöntem, geleneksel transfer öğrenme yöntemlerine göre çok daha az eğitim verisiyle bile etkili sonuçlar verir ve sıfır örnekli öğrenme yeteneklerine sahiptir. TARS, farklı metin sınıflandırma görevlerinde (duygu analizi, konu tespiti vb.) test edilmiştir. Bu testler, TARS'ın az örnekli ve sıfır örnekli senaryolarda geleneksel yöntemlere göre önemli ölçüde daha iyi performans gösterdiğini ortaya koymuştur (Halder, Kishaloy, *et al.* 2020).

TARS, doğal dil çıkarımından ilham alarak, giriş metnini etiketlerle birleştirir ve etiket doğruysa "Doğru", aksi takdirde "Yanlış" olarak tahmin etmektedir (Ziletti, Angelo, *et al.* 2022). TARS modeli ön-eğitilmiş bir BERT modelidir. Önemli bir özelliği modelin encoder ve decoder arasında paylaşılabilmesi ve aktarılabilmesidir. Orijinal TARS modeli her bir eğitim örneğinin en az bir pozitif (1) etikete atandığı varsayımına dayanmaktadır (Martinez-Gomez *et al.* 2021). Halder ve arkadaşlarının 2020'de geliştirdiği TARS modeli, yeni bir görevin önceki göreve önemli ölçüde benzemediği durumlarda, bu yeni görev için modelin baştan eğitilmesini gerektirmektedir. Öte yandan, yeni görev önceki göreve benzer nitelikteyse, BERT tabanlı ince ayar yöntemi uygulanarak daha önceki görevlerden kazanılan ağırlıklar, yeni görev için başlangıç noktası olarak kullanılmaktadır. TARS modeline ait mimari Şekil 5'te verilmiştir.



**Şekil 5.** TARS modeli mimarisi (Halder, Kishaloy, *et al.* 2020)

TARS, az örnekli ve sıfır örnekli ayarlarda güçlü sonuçlar göstermektedir. Fakat tüm bunların yanında karşılaşılan bazı sınırlamalarda bulunmaktadır. Çok sayıda sınıf içeren durumlarda zorluklarla karşılaşılmaktadır. Her bir etiket adayı için modelin ayrı ayrı çalıştırılması gerektiğinden, çok fazla sınıf olduğunda bu işlem zaman ve kaynak açısından zorlayıcı olmaktadır. Ayrıca, modelin olumsuz etiket örnekleme yöntemi, büyük ölçekli sınıflandırmalarda etkili olmayabilmektedir. Çünkü model, gerçekte zor olmayan etiketleri olumsuz örnek olarak algılayarak yanıltılmaktadır. Bu da öğrenme sürecini olumsuz etkilemektedir.

Özetle, TARS, metin sınıflandırmasında, özellikle az veriyle çalışırken veya tamamen yeni sınıflarla karşılaştığında, geleneksel yöntemlerin karşılaştığı zorlukların üstesinden gelmek için tasarlanmış yenilikçi bir yaklaşımdır. Bu yöntem, metin ile sınıf etiketlerinin anlamsal bilgilerini birleştirerek daha etkili ve esnek bir sınıflandırma modeli sunmaktadır.

## Literatür Taraması

Hughes ve Palen (2009) çalışmalarında, sosyal medyanın afet zamanı halk iletişimi için nasıl bir platform haline geldiğini incelemişlerdir. Bu, bireylerin afet anında nasıl tepki verdiklerini, hangi bilgileri paylaştıklarını ve bu bilgilerin sahadan nasıl toplandığını anlamak için kritik bir çalışmadır.

Sakaki, Okazaki ve Matsuo'nun (2010) yaptığı çalışma, Twitter'ın depremleri gerçek zamanlı olarak algılama yeteneğini vurgulamaktadır. Araştırmacılar, spesifik tweet

aktivitelerini izleyerek depremlerin hemen ardından bile gerçekleşen depremlerin tespit edilebileceğini belirtmişlerdir.

Öte yandan, deprem verilerini analiz eden bir başka çalışma da Avvenuti ve arkadaşları (2014) tarafından gerçekleştirilmiştir. Araştırmalarında, İtalya’da depremler sırasında ve sonrasında toplanan Twitter verilerini kullanarak, bu verilerin hasar raporlama ve değerlendirme süreçlerinde nasıl etkili bir araç olabileceğini incelemişlerdir.

Muhammad Imran ve arkadaşları (2014), 2013 ve 2015 yılları arasında gerçekleşen 19 farklı kriz sırasında toplanan insan tarafından etiketlenmiş Twitter verisi oluşturmuşlardır. Etiketlemelerin faydasını göstermek için makine öğrenimi sınıflandırıcıları eğitilmiştir. 52 milyon krizle ilgili tweet üzerinde eğitilmiş ilk en büyük word2vec kelime gömütlerini yayınlamışlardır. Kriz durumunun aniden başlaması sırasında, etkilenen insanlar, zamanında ve etkili bir şekilde işlenirse durumsal farkındalık ve diğer insani afet yanıt çabaları için kullanılabileceği incelemişlerdir.

Benzer şekilde Imran ve arkadaşları (2013) yaptıkları çalışmada, sosyal medya platformlarından afetle ilgili bilgilerin nasıl etkin bir şekilde çıkarılabileceğini ele almaktadır. Çalışma, sosyal medyadan elde edilen verilerin, afet durumlarında kritik bilgileri sağlama potansiyeline sahip olduğunu vurgulamaktadır. Afet durumlarında, insanlar genellikle sosyal medyayı bilgi paylaşımı ve yardım talebi için kullanmaktadır. Bu çalışmada, verilerin nasıl toplanacağını, işleneceğini ve analiz edileceğini inceleyerek, afet yanıtı ve yönetimi için pratik çözümler sunmayı amaçlamaktadır.

Tien Nguyen ve arkadaşlarının 2016 tarihli çalışması, sosyal ağlardaki krizle ilgili verilerin hızlı sınıflandırılmasını ele almaktadır. Bu çalışma, özellikle Evrişimsel Sinir Ağları kullanarak sosyal medya verilerinin kriz durumlarında nasıl etkin bir şekilde analiz edilebileceğine odaklanmıştır. Çalışma, sosyal ağlardaki büyük veri akışının, doğal afetler veya acil durumlar gibi kriz anlarında önemli bilgiler içerebileceğini vurgulamaktadır. Araştırmacılar, CNN teknolojisini kullanarak bu verilerin hızla ve etkin bir şekilde sınıflandırılmasının mümkün olduğunu göstermişlerdir. Bu sınıflandırma, kriz yanıt ekiplerine gerçek zamanlı bilgi sağlayarak hızlı ve etkili müdahalelere olanak tanıyabileceğini savunmaktadır.

Salvador Medina Maza ve arkadaşları (2020) yaptıkları çalışmada sosyal medyanın kriz durumlarında bilgi paylaşımında ne kadar kritik bir role sahip olduğunu, ancak, bu veri akışlarının gerçek zamanlı olarak analizi, önemli bilgilerin geniş veri yığınları arasından seçilmesiyle ilgili zorlukları da beraberinde getirdiğinden bahsetmektedir. Mevcut modellerin çoğu, daha önceki olaylardan elde edilen verilere dayanarak eğitilmiş olup, bu da olayın

lokasyonu gibi özel önyargıları modele taşıyabileceğini ve bu tür önyargıların, yeni ve görülmemiş olaylarda modelin geliştirilmesini ve performansını olumsuz etkileyeceğini savunmuşlardır. Bu bağlamda önyargıları minimize etmeyi amaçlayan rakip bir nöral model eğitmişlerdir. Bu model, Twitter mesajlarının önemini daha doğru bir şekilde sınıflandırmayı hedeflemektedir.

F. Ofli, F. Alam ve M. Imran'ın 2020 tarihli çalışmasında, özellikle çok modlu derin öğrenme tekniklerini kullanarak afet yanıtı için sosyal medya verilerinin analizini ele almaktadır. Çalışmada, sosyal medya platformlarından elde edilen verilerin, afet durumlarına yanıt vermede önemli bir kaynak olduğu vurgulanmaktadır. Bu veriler genellikle metin, görüntü, video ve ses gibi çeşitli formatlarda bulunmaktadır. Çalışmada, bu farklı veri türlerini entegre ederek daha etkili bir afet yanıtı analizi yapılması için çok modlu derin öğrenme modelleri kullanılmaktadır. Çok modlu derin öğrenme, farklı veri türlerini (örneğin metin ve görüntü) birleştirerek bu verilerden daha kapsamlı ve anlamlı bilgiler çıkarılmasını sağlamaktadır. Bu yaklaşım, özellikle karmaşık ve dinamik afet durumlarında, veri setlerinin daha derinlemesine analiz edilmesine ve daha doğru sonuçlar elde edilmesine olanak tanımaktadır. Sonuç olarak, çalışma, çok modlu derin öğrenme modellerinin, sosyal medya verilerini afet yanıtı bağlamında analiz etmede etkili araçlar olduğunu savunmaktadır.

Aldamen ve arkadaşları (2023) yaptıkları çalışmayla 6 Şubat 2023'teki Türkiye depremin de Twitter'ın kriz iletişimindeki rolünün nasıl evrildiğini incelemektedir. Bu evrimi anlamak için, 1999 İzmit ve 2000 İzmir depremlerindeki Twitter kullanımına bakılmıştır. Araştırma, Twitter'ın zamanla sadece bilgi toplama aracı olmaktan çıkıp, aktif bir şekilde afetzedelerin kurtarılmasına yardımcı olan bir platforma dönüştüğünü göstermektedir. Özellikle deprem sırasında kaybolanları bulma ve anlık durum güncellemeleri konusunda Twitter'ın kritik bir role sahip olduğunu, bu da medyanın gözetim görevini yerine getirdiğini teyit etmektedir.

Toraman ve arkadaşları (2023) ise yaptıkları çalışmada 6 Şubat 2023 Türkiye ve Suriye depremindeki enkaz altında kalan insanların sosyal medya üzerinden bildirilmesi durumunu göz önüne alarak bir sınıflandırma çalışması yapmıştır. Bu çalışmada enkaz altında bir depremzedenin olduğunda ve bir adres içeren Twitter mesajları sınıflandırılmış ve haritada işaretlenmiştir.

Romain Meunier ve arkadaşları (2023) ise yaptıkları çalışma ile, kriz yönetimi için manuel olarak etiketlenmiş yeni bir Fransızca veri setini sunmaktadırlar. Krizler sırasında kurtarma hizmetlerine en iyi şekilde yardımcı olmak için Twitter mesajlarını ilgili olduklarına,

aciliyetlerine ve taşıdıkları niyete göre sınıflandırarak çeşitli otomatik öğrenme modellerini test etmişlerdir.

U. Şahin ve arkadaşlarının 2023 tarihli çalışması, sosyal medya mesajlarında, 6 Şubat 2023 Türkiye ve Suriye deprem felaketiyle ilişkili olanlarda, nefret söyleminin tespiti üzerine odaklanmaktadır. Bu araştırma, sınırlı veya hiç etiketlenmiş veriyle çalışabilen 'zero-shot' ve 'few-shot' öğrenme yöntemlerini kullanarak nefret söylemi tespiti yapmayı amaçlamaktadır. Bu yaklaşımlar, özellikle veri kısıtlı durumlar için oldukça faydalıdır ve etiketleme maliyetlerini azaltmaya yardımcı olmaktadır. Çalışma, deprem gibi felaket durumlarında sosyal medyada ortaya çıkan nefret söylemi mesajlarının otomatik olarak tespit edilmesi için bu yöntemlerin nasıl kullanıldığını incelemektedir. Bu tür bir analiz, deprem sonrası sosyal medya ortamında nefret söyleminin izlenmesi ve mücadele edilmesi açısından önemlidir. Araştırmacılar, çeşitli derin öğrenme ve doğal dil işleme tekniklerini kullanarak, sınırlı veriyle bile etkili nefret söylemi tespiti yapabilecek modeller geliştirmeye odaklanmışlardır. Bu çalışma, sosyal medya analizi ve afet yönetimi alanlarında önemli bir katkı sunarak, kriz durumlarında nefret söyleminin tespit edilmesi ve yönetilmesi için yeni yaklaşımlar sağlamaktadır.

## MATERYAL VE METOT

Çalışma kapsamında deprem sonrası atılan Twitter mesajlarının toplanıp ihtiyaçlar doğrultusunda sınıflandırılması işlemi gerçekleştirilmiştir. Bu işlem için hazır bulunan bir veri seti üzerinden ilerlenmiştir. Bu veri setinden çalışma kapsamı doğrultusunda 2000 tane Türkçe Twitter verisi seçilmiş ve seçilen hedefler doğrultusunda etiketlenmiştir. Etiketlenen bu veri setinin öncelikle tamamını kullanarak BERT, RoBERTa ve DistilBERT gibi az etiketli veri ile iyi sonuçlar veren modeller kullanılarak sınıflandırma işlemleri yapılmıştır. Ardından geleneksel makine öğrenmesi yöntemleri ile kıyaslama yapılması amaçlanarak KNN, Destek Vektör Makinesi (SVM), Naive Bayes, Karar Ağacı ve Lojistik Regresyon gibi sınıflandırma algoritmaları da çalışmada yer almaktadır. Tüm bunların yanı sıra zero-shot ve few-shot gibi uygulamalara da yer verilmiştir. Çalışma kapsamında kullanılan yöntemlerin uygulanması tez içerisinde yöntem başlığı altında detaylı bir şekilde açıklanmıştır.

### Veri Seti

Çalışmanın hedefi doğrultusunda 6 Şubat 2023 depremine ait Twitter verilerine ihtiyaç duyulmuştur. Bu ihtiyacı karşılamak için Twitter'ın sunduğu API'ler mevcuttur. Fakat bu çalışma için Kaggle'da bu deprem ile ilgili toplanmış 478.052 verinin olduğu "Turkey and Syria Earthquake Tweets" adlı bir veri seti ele alınmıştır. Bu veri seti depreme ait 17 dilde tweet içermektedir. Veri setinde tweete ait farklı birçok bilgi bulunmaktadır. Veri setine ait sütunlar ve içerdiği bilgiler Tablo 2'de verilmiştir.

**Tablo 2.** Veri Setindeki Twitter Mesajlarına Ait Bilgiler

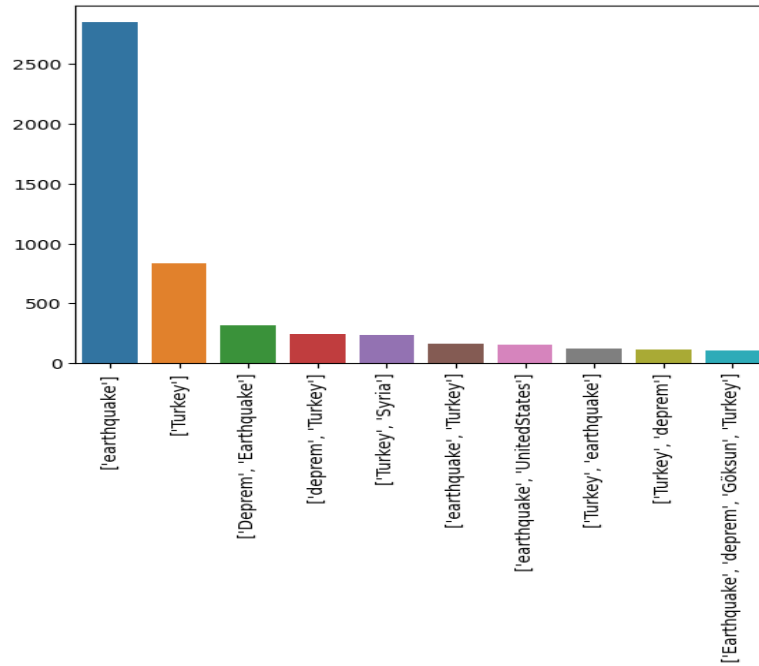
| Özellik         | İçeriği                                                       |
|-----------------|---------------------------------------------------------------|
| Date            | Tweetin atıldığı tarih ve saat                                |
| Content         | Tweetin içeriği                                               |
| Hashtags        | Tweette bulunan hashtag'ler                                   |
| Like_Count      | Tweetin beğeni sayısı                                         |
| Rt_Count        | Tweetin kaç kez retweet edildiği sayısı                       |
| Followers_Count | Tweeti atan hesabın takipçi sayısı                            |
| Isverified      | Takipçi sayısına göre hesabın doğruluğunu gösteren bir etiket |
| Language        | Tweetin yazıldığı dil                                         |
| Coordinates     | Tweetin yazıldığı yerin koordinatları                         |
| Place           | Tweetin yazıldığı yer                                         |
| Source          | Tweetin atıldığı kaynak (Android, IOS, web APP vs.)           |

Veri setinde 17 farklı dil için tweet bulunmaktadır. Fakat diller arası dağılım eşit değildir. Veri setinin dillere göre dağılımı Tablo 3’te gösterilmektedir.

**Tablo 3.** Veri Setinin Dillere Göre Dağılımı (Sadece 5 dil için verilmiştir.)

| Dil        | Tweet Sayısı |
|------------|--------------|
| İngilizce  | 9.509        |
| Türkçe     | 5.517        |
| Arapça     | 563          |
| Sırpça     | 588          |
| İspanyolca | 458          |

Twitter API aracılığı ile mesajlar alınırken hashtag’ler anahtar kelime olarak kullanılmıştır. Belirtilen tarihler aralığında belirli konu ile ilgili mesajlar alınabilmektedir. Bu veri seti hazırlanırken kullanılmış olan hashtagler ve dağılımları Şekil 5’te verilmiştir.



**Şekil 6.** Twitter API’den twitter mesajları alınırken kullanılan hashtag dağılımı

Çalışma kapsamında Türkçe Twitter mesajları üzerinde çalışma yapılacağı için bu veri seti içerisinde sadece Türkçe diline ait olanlar seçilmiştir. Sonuç olarak 5.517 Türkçe tweet verisinin olduğu bir veri seti elde edildi. Bu veri setinde çalışmanın amacı doğrultusunda Twitter mesajlarının etiketlenmesi ihtiyacı ortaya çıkmıştır. Bu ihtiyaç doğrultusunda seçilen 2000 veri etiketleme işlemine alındı.

Etiketleme işlemi manuel olarak yapılmış olup, bu işlem 3 farklı kişi tarafından yapılmıştır. 3 farklı kişi tarafından yapılan etiketleme işlemi sonrasında Fleiss’ Kappa değeri hesaplanmıştır. Fleiss’ Kappa değeri, üç veya daha fazla değerlendiricinin bir dizi ögeye verdiği

derecelendirmelerin tutarlılığını ölçmek için kullanılan bir istatistiksel ölçümdür. Genellikle, araştırmacılar veya uzmanlar tarafından yapılan değerlendirmelerdeki anlaşma düzeyini belirlemek için kullanılmaktadır. Kappa değeri, sadece iki değerlendiricinin anlaşmasını değerlendirmek için kullanılan Cohen's Kappa'nın genelleştirilmiş bir formudur. Fleiss' Kappa, şans eseri anlaşma düzeyini hesaba katarak gerçek anlaşma düzeyini daha doğru bir şekilde yansıtmaktadır. Bu ölçüm, değerlendirmeler arasındaki anlaşma düzeyini -1 ile +1 arasında bir değerle ifade eder. +1 mükemmel anlaşmayı, 0 şans eseri anlaşmayı ve -1 tamamen uyumsuzluğu temsil etmektedir. Yapılan bu hesaplamalar sonucunda Fleiss' Kappa değeri %89,43 olarak elde edilmiştir. Bu değer +1'e oldukça yakın olduğundan değerlendiriciler arasında mükemmel uyum olduğunu göstermektedir. Son olarak farklı olarak etiketlenen veriler için tartışılarak fikir birliği sağlanmış ve 2000 adet 3 etiket ile etiketlenmiş veri seti elde edilmiştir. Fleiss' Kappa değerinin hesaplanmasında kullanılan matematiksel gösterim verilmiştir. Po, gözlenen anlaşmayı Pe ise beklenen anlaşmayı ifade etmektedir. (Joseph L. *et al.* 2003)

$$kappa(\kappa) = \frac{P_o - P_e}{1 - P_e} \quad (1)$$

Etiketler seçilmeden önce Twitter mesajları genel olarak incelenmiş ve doğru sınıflar amaç ile orantılı bir şekilde seçilmiştir. Seçilen etiketler Tablo 4'te ve etiketlere ait veri sayıları ise Tablo 5'te detaylı olarak verilmiştir.

**Tablo 4.** Seçilen Etiketler ve Açıklaması

| Etiket_Adı     | Etiket_Id | Anlamı                                                                                                                                           |
|----------------|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ilgisiz</b> | 0         | Kişinin duygularından bahsettiği veya siyasi yorumlar yaptığı mesajlar                                                                           |
| <b>enkaz</b>   | 1         | Enkaz altında kalan bir kişinin yardım istediği veya enkazdan ses geldiği için ekiplerden yardım istenen mesajlar (adres içeriği mevcut)         |
| <b>ihtiyaç</b> | 2         | Depremzedeler için ihtiyaçların belirtildiği, barınma noktalarının bildirildiği veya erzak dağıtım/toplama alanlarının bilgisini içeren mesajlar |

**Tablo 5.** Örneklerin Kategoriler Arası Dağılımı

| Etiket Adı     | Veri Sayısı |
|----------------|-------------|
| <b>ilgisiz</b> | 693 tane    |
| <b>enkaz</b>   | 654 tane    |
| <b>İhtiyaç</b> | 653 tane    |

Elde edilen bu etiketli veri setini seçilen tüm modeller doğru ve verimli bir şekilde kullanabilmek için gürültülerin giderilmesi ihtiyacı doğmuştur. Bu ihtiyaç doğrultusunda veri seti incelenerek gerekli ön işlem adımları belirlenmiştir. Veri seti Twitter verilerinden oluştuğu için içeriğinde; gereksiz noktalama işaretleri, emoji, internet bağlantıları, hashtagler, kullanıcı hesap adları gibi gürültülerin olduğu tespit edilmiştir. Bu doğrultuda veri setinde tweetin içerisinde “#” ile başlayan hashtagler ve “@” ile başlayan hesap etiketleri kaldırılmıştır. Ardından internet bağlantıları, nokta (.) hariç diğer noktalama işaretleri ve emoji kaldırılarak tweet metinleri daha temiz ve anlaşılır hale getirilmiştir. (Nokta ve sayısal ifadeler kaldırılmamıştır. Çünkü enkaz olarak etiketlenen Twitter mesajlarında özellikle adres kısımlarında nokta bulunmaktadır. Kaldırıldıklarında adres bilgisi bozulmaktadır. İleride yapılacak çalışmalar düşünüldüğünde adres bilgisinin korunmasında fayda olduğu düşünülmüştür. Ön işlem adımları Kuramsal Temeller bölümünde daha detaylı olarak gösterilmiştir.)

Etiketli veri setinde ön işleme adımlarından öncesi ve sonrası Twitter mesajlarının içeriği Şekil 6, Şekil 7, Şekil 8, ve Şekil 9, Şekil 10, Şekil 11 ve Şekil 12 ‘de gösterilmiştir.

|       |   |                                                                                                                                                                                               |
|-------|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| enkaz | 1 | @BabalaTV General şükürü kanatlı mahallesi Fevzi Paşa caddesi İncesu apartmanı daire 3 Antalya/Hatay #Turkey #DEPREMOLDU #hatayyardimbekliyor #HatayKaderineTerkEtme #AFAD #a                 |
| enkaz | 1 | GÜLBAHAR TEKE Gaziantep Yardım çağrısında ismi geçer veya görürseniz lütfen ulaşın bana! Instagram: busraxyw_06 #sondakikadeprem #Turkey #Turkey #PrayForTurkey #Antep #antepi                |
| enkaz | 1 | 🔥 MARAŞ PENTA PARK 3 ÇOK ACIL THERMAL KAMERA İHTİYAÇLARI VAR 🔥 YAYALIM ARKADAŞLAR LÜTFEN #sesimiduyanvarmi #deprem #Turkey #PrayForTurkey #HelpTurkey                                         |
| enkaz | 1 | Hala bu binalarda insanlar var lütfen yardımcı olun LÜTFEN PAYLAŞIN RTLEYİN #seferberlik #Turkey #TurkeyEarthquake #tahadiyalmaz #acilvinc #fuliaozturk #DEPREM #ENKAZ https://t.co           |
| enkaz | 1 | @lordsinov General şükürü kanatlı mahallesi Fevzi Paşa caddesi İncesu apartmanı daire 3 Antalya/Hatay #Turkey #DEPREMOLDU #hatayyardimbekliyor #HatayKaderineTerkEtme #AFAD #a                |
| enkaz | 1 | Kahramanmaraş Göksun İlçesinin köylerine hala yardım gitmiş değil. Kanlıkavak, köyleri, kavşut... İnsanlar ya soğuktan ya da açlıktan ölmek üzereler.. Biran önce yardım gitmeli. #TurkeyEar  |
| enkaz | 1 | Sümerevler mahallesi bobbey caddesi ADIYAMAN 8KİŞİ VER 5'inden hala ses geliyor lütfen yardım edin @AFADBaskanlık @BabalaTV @haluklevent @OguzhanUgur @KendineMuzisyen @                      |
| enkaz | 1 | GAZİANTEP ! AYŞE MEHMET POLAT SİTESİ ! A BLOK ! 3. 4. KATLARDAN BİRDEN FAZLA KİŞİNİN SESİ DUYULUYOR ACIL BİLGİSİ OLAN YAZSSIN #sondakikadeprem #gaziantep #Turkey #                           |
| enkaz | 1 | Battalgazi, Dabakhane Mah. No; Turgut Temelli Cad. No:26, 44100 MALATYA KIRÇUVAL OTEL YARDIMIM NE OLURSUNUZ YARDIM #Turkey #TurkeyEarthquake #sondakikadeprem #PrayForT                       |
| enkaz | 1 | 🔴 Karapınar mahallesi, Yürüyüş yolu caddesi Merkez Adıyaman. 5 katlı bina çöktü İçerde Mehmet üder eşi Dilek Üder ve çocukları Derin ve Deniz Üder göçük altında. @AFADBaskanlık @A           |
| enkaz | 1 | !!ACİL!! Duran Oğuz Temizkan ve annesi yaklaşık 48 saatır soğuk havada enkaz altında. Aynı adreste 18 aylık bir bebeğin olduğu söyleniyor. Akevler Mahallesi, 432. Sokak Saray Apt. No:7 D.   |
| enkaz | 1 | ! ACİL ACİL ! arkadaşım ve ailesinden haber yok lütfen yardım edin lütfen olmesinler Ayca Doğruer 05075011078 Hatay/Antakya Odabaşı Mahallesi, Uğur Mumcu Caddesi, Mahmut K                   |
| enkaz | 1 | ismail ve zaziye korkmaz. Odabaşı Mahallesi Uğur Mumcu caddesi rende sitesi -Hatay. arkadaşımın kuzeni ve hamile acilen yayar mısınız lütfen. kardeşinin numarası 05457763279. #seferber      |
| enkaz | 1 | Hatay/Antakya Akademi hastanesi arkası akademi apartmanı yıldız ev yemekleri ev üstü 5 kat. Öznur azazi hasan göçmen layla göçmen Öznur 45 layla 21 hasan 17 yaş #sondakikadeprem #           |
| enkaz | 1 | ADIYAMAN MERKEZ CUMHURİYET MAHALLESİ 2596.SOKAK HALI SAHA KARŞISI 4 KATLI BİNA ENKAZ ALTINDA KİŞİLER VAR İLK 45 SAATİ AŞKIN YARDIM GİTMİYOR LÜTFEN 🇹🇷 #adiyamanede                            |
| enkaz | 1 | Üniversiteden arkadaşımız, güzel Ebrar'ımıza ve ailesine ulaşamıyoruz. Yeni Şehir Mahallesi Zübeyde Hanım Bulvarı Işık Apartmanı 35 Dulkadiroğlu/KAHRAMANMARAŞ #SONDAKİDEPREM #               |
| enkaz | 1 | İstasyon mahallesi cengizhan sokak no/16-17 K.maraş Türkoğlu İki gündür enkaz altındalar 5 kişilik bir aile bugün sesler kesilmiş ve hiçbir şekilde yanıt alınmıyor lütfen yardım edin #depre |

Şekil 7. Enkaz olarak etiketli twitter mesajlarının ön işleme öncesi durumu

|       |   |                                                                                                                                                                                                                             |
|-------|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| enkaz | 1 | general şükürü kanatlı mahallesi fevzi paşa caddesi İncesu apartmanı daire 3 antakya/hatay                                                                                                                                  |
| enkaz | 1 | gülbahar teke gaziantep yardım çağrısında ismi geçer veya görürseniz lütfen ulaşın bana! instagram busraxyw06                                                                                                               |
| enkaz | 1 | maraş penta park 3 çok acil thermal kamera ihtiyacı var yayalim arkadaşlar lütfen                                                                                                                                           |
| enkaz | 1 | hala bu binalarda insanlar var lütfen yardımcı olun lütfen paylaşın rtleyin                                                                                                                                                 |
| enkaz | 1 | general şükürü kanatlı mahallesi fevzi paşa caddesi İncesu apartmanı daire 3 antakya/hatay                                                                                                                                  |
| enkaz | 1 | kahramanmaraş göksun ilçesinin köylerine hala yardım gitmiş değil. kanlıkavak köyleri kavşut... insanlar ya soğuktan ya da açlıktan ölmek üzereler.. biran önce yardım gitmeli.                                             |
| enkaz | 1 | sümerevler mahallesi bobbey caddesi adıyaman 8kişi ver 5'inden hala ses geliyor lütfen yardım edin                                                                                                                          |
| enkaz | 1 | gaziantep ayşe mehmet polat sitesi a blok 3. 4. katlardan birden fazla kişinin sesi duyuluyor acil bilgisi olan yazssın                                                                                                     |
| enkaz | 1 | battalgazi dabakhane mah. no turgut temelli cad. no26 44100 malatya kırçuval otel yardımım ne olursunuz yardım                                                                                                              |
| enkaz | 1 | karapınar mahallesi yürüyüş yolu caddesi merkez adıyaman. 5 katlı bina çöktü İçerde mehmet üder eşi dilek üder ve çocukları derin ve deniz üder göçük altında.                                                              |
| enkaz | 1 | acil duran oğuz temizkan ve annesi yaklaşık 48 saatır soğuk havada enkaz altında. aynı adreste 18 aylık bir bebeğin olduğu söyleniyor. akevler mahallesi 432. sokak saray apt. no                                           |
| enkaz | 1 | acil acil arkadaşım ve ailesinden haber yok lütfen yardım edin lütfen olmesinler ayca doğruer 05075011078 hatay/antakya odabaşı mahallesi uğur mumcu caddesi mahmut kar                                                     |
| enkaz | 1 | ismail ve zaziye korkmaz. odabaşı mahallesi uğur mumcu caddesi rende sitesi hatay. arkadaşımın kuzeni ve hamile acilen yayar mısınız lütfen. kardeşinin numarası 05457763279                                                |
| enkaz | 1 | hatay/antakya akademi hastanesi arkası akademi apartmanı yıldız ev yemekleri ev üstü 5 kat. öznur azazi hasan göçmen layla göçmen öznur 45 layla 21 hasan 17 yaş                                                            |
| enkaz | 1 | bir bebek açlıktan ölmek üzere!!! araç plakası:02 BA 359 ADIYAMAN GÖLBAŞI-CUMHURİYET MAHALLESİ-KARAYILAN CADDESİ ÜZERİNDİR. Tel No:+90 542 496 3341/İsim: Rahime #sesimiduyanvarmi #deprem #Turkey #PrayForTur              |
| enkaz | 1 | Hatay Antakya Köprübaşı YARDIM DAĞITILMAYA BAŞLANMIŞ!!! #DEPREMOLDU #sondakikadeprem #deprematay #TurkeyEarthquake #Turkey @DepremDairesi @BabalaTV https://t.co/kehtPMLueX                                                 |
| enkaz | 1 | 🔥🔥🔥🔥🔥 ÇOK ACİL!!! Bir bebek açlıktan ölmek üzere, yakın olan varsa yardım edebilir mi? Araç plakası 02 BA 359 Adıyaman Gölbaşı cumhuriyet mahallesi karayılan caddesindedir. Rahime/Tel: 0542 496 33 41 #Turkey             |
| enkaz | 1 | Hatay'da olup aracı olan biri lazım, kadın hamile zorlu köyünde kanaması başlamış. Havaalanına yetmesi gerek. Durumu çok çk kötü +905444643397 EBRU DOĞAN #hatayyardimbekliyor #Hatay #TurkeyQuake #Turkey #sondak          |
| enkaz | 1 | Yarın için deprem bölgesinden ANTALYA üzerinden dönüş yapacak yardım aracı var mıdır? Depremde arkadaşım için Bursa kadar yardımcı olabilecek bir araç arıyoruz. #Antalya #sondakikadeprem #DEPREMOLDU #Turkey #afa         |
| enkaz | 1 | Bu site, depreme ilgili bağış, barınma, gerekli IBAN numaraları, geçici sinyal, deprem anı ve sonrasında gerekli yazılara kolaylıkla ulaşabilmeyi sağlıyor lütfen yayalım. #MARAS #DiYarbakırdeprem #adanadeprem #Turkey #h |

Şekil 8. Enkaz olarak etiketli twitter mesajlarının ön işleme sonrası durumu

|         |   |                                                                                                                                                                                                                             |
|---------|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| İhtiyaç | 2 | 50 kişilik itfaiye ekibi Maraş havalimanında kalmış. Hatay'a gidecek araba arıyorlar. İletişim:0532 704 98 62 Muhammed Ataş #DEPREMOLDU #seferberlik #Turkey #Turkey #hatayyardimbekliyor #Elbistan #YARDIMEDİN #yardim     |
| İhtiyaç | 2 | !!! ADIYAMAN HAVALIMANI !!! İZDİHAM VAR ARKADAŞIMIZ HAVALIMANINDA İZDİHAMIN İÇİNDE ERZAKLARI YOK LÜTFEN YARDIM EDİN LÜTFEN BİR ŞEY YAPIN LÜTFEN #Turkey #TurkeyEarthquake #ACILYARDIM #A                                    |
| İhtiyaç | 2 | Lütfen İsaat işçileri gönderilsin bu işi onlar daha iyi biliyorlar... #deprematay #Turkey                                                                                                                                   |
| İhtiyaç | 2 | Tüm operatörler deprem bölgesinde olan herkese ücretsiz dakika ve internet yüklemeli!! İnsanların sesi olun! @Turkcell @VodafoneTR @TurkTelekom #Turkey #PrayForTurkey #DEPREMOLDU #depreme #deprem                         |
| İhtiyaç | 2 | Elbistan/Kahramanmaraş arkadaşım cadır ismine gibi yardımların hala ulaşmadığını söyledi çok zor durumdadır esentepe mahallesi camiler sokağında parkta lütfen paylaşın lütfen @haluklevent @OguzhanUgur @AFADasi           |
| İhtiyaç | 2 | BİR BEBEK AÇLIKTAN ÖLMEK ÜZERİ!!! Araç plakası:02 BA 359 ADIYAMAN GÖLBAŞI-CUMHURİYET MAHALLESİ-KARAYILAN CADDESİ ÜZERİNDİR. Tel No:+90 542 496 3341/İsim: Rahime #sesimiduyanvarmi #deprem #Turkey #PrayForTur              |
| İhtiyaç | 2 | Hatay Antakya Köprübaşı YARDIM DAĞITILMAYA BAŞLANMIŞ!!! #DEPREMOLDU #sondakikadeprem #deprematay #TurkeyEarthquake #Turkey @DepremDairesi @BabalaTV https://t.co/kehtPMLueX                                                 |
| İhtiyaç | 2 | 🔥🔥🔥🔥🔥 ÇOK ACİL!!! Bir bebek açlıktan ölmek üzere, yakın olan varsa yardım edebilir mi? Araç plakası 02 BA 359 Adıyaman Gölbaşı cumhuriyet mahallesi karayılan caddesindedir. Rahime/Tel: 0542 496 33 41 #Turkey             |
| İhtiyaç | 2 | Hatay'da olup aracı olan biri lazım, kadın hamile zorlu köyünde kanaması başlamış. Havaalanına yetmesi gerek. Durumu çok çk kötü +905444643397 EBRU DOĞAN #hatayyardimbekliyor #Hatay #TurkeyQuake #Turkey #sondak          |
| İhtiyaç | 2 | Yarın için deprem bölgesinden ANTALYA üzerinden dönüş yapacak yardım aracı var mıdır? Depremde arkadaşım için Bursa kadar yardımcı olabilecek bir araç arıyoruz. #Antalya #sondakikadeprem #DEPREMOLDU #Turkey #afa         |
| İhtiyaç | 2 | Bu site, depreme ilgili bağış, barınma, gerekli IBAN numaraları, geçici sinyal, deprem anı ve sonrasında gerekli yazılara kolaylıkla ulaşabilmeyi sağlıyor lütfen yayalım. #MARAS #DiYarbakırdeprem #adanadeprem #Turkey #h |

Şekil 9. İhtiyaç olarak etiketli twitter mesajlarının ön işleme öncesi durumu

|         |   |                                                                                                                                                                             |
|---------|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| İhtiyaç | 2 | 50 kişilik itfaiye ekibi maraş havalimanında kalmış, hataya gidecek araba arıyorlar. İletişim0532 704 98 62 muhammed ataş n                                                 |
| İhtiyaç | 2 | adiyaman havalimanizidham var arkadaşımız havalimanında izdihamin içinde erzakları yok lütfen yardım edin lütfen bir şey yapınılufen                                        |
| İhtiyaç | 2 | lütfe inşaat işçileri gönderilsin bu işi onlar daha iyi biliyorlar...                                                                                                       |
| İhtiyaç | 2 | tüm operatörler deprem bölgesinde olan herkese ücretsiz dakika ve internet yüklemeli insanların sesi olun                                                                   |
| İhtiyaç | 2 | elbistankahramanmaraş arkadaşım çadır ısına gibi yardımların hala ulaşmadığını söyledi çok zor durumdalar esentepe mahallesi camiler sokağındaki parktalar lütfen paylaşıp  |
| İhtiyaç | 2 | bir bebek ağıltan ölmek üzere araç plakası02 ba 359 adiyaman gölbaşicumhuriyet mahallesikarayılan caddesi üzerindeler. tel no90 542 496 3341isim rahime                     |
| İhtiyaç | 2 | hatay antakya köprübaşında yardım dağıtılmaya başlanmış                                                                                                                     |
| İhtiyaç | 2 | çok acil bir bebek ağıltan ölmek üzere yakın olan varsa yardım edebilir mi araç plakası 02 ba 359 adiyaman gölbaşı cumhuriyet mahallesi karayılan caddesindeler rahimetel t |
| İhtiyaç | 2 | hatayda olup aracı olan biri lazım kadın hamile orlu köyünde kanaması başlamış. durumu çok kötü 905444643397 ebru doğan                                                     |
| İhtiyaç | 2 | yarın için deprem bölgesinden antalya üzerinden dönüş yapacak yardım aracı var mıdır depremde arkadaşım için bursa kadar yardımcı olabilecek bir araç arıyoruz              |
| İhtiyaç | 2 | bu site depremle ilgili bağı barınma gerekli iban numaraları geçici sığınaklar deprem anı ve sonrasında gerekli yazılara kolaylıkla ulaşabilmeyi sağlıyor lütfen yayalım.   |

**Şekil 10.** İhtiyaç olarak etiketli twitter mesajlarının ön işleme sonrası durumu

|         |   |                                                                                                                                                                                                                                       |
|---------|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| İlgisiz | 0 | Daha kötüsünü yaşamadım. Kahramanmaraşta 7.4 büyüklüğünde bir deprem. Enkaz altında olanlar var Allah'ım acil yardım et o insanlara. sen beterinden kuru. Bir daha olmaz inşallah. #deprem #batman #Kahramanmaraş #ear                |
| İlgisiz | 0 | Arkadaşlar herkes adresler paylaşıyor rt ediyor bunlar doğru ama inanılmaz derecede yıkılan enkaz altında kalan bina var gibi oyuşten tüm illerden ekipler getirilmeli büyük çadırlar kurulmalı tüm türkiye seferberlik olmalı #depre |
| İlgisiz | 0 | Kayseride depremin etkisi. Çevre illerdeki etkisinin yayılması adına bir daha paylaşıyorum. cümleten geçmiş olsun. #deprem #DEPREMOLDU #earthquake #Kandilli #Kahramanmaraş #sallandık #sanlıurfa #kayseri https://t.co/DT            |
| İlgisiz | 0 | Lütfen gerekmedikçe hatları meşgul etmeyin yardıma ihtiyacı olan insanlar var duyarlı olun biraz 7.8 şiddetinde deprem 🇹🇷 #Gaziantep #Kahramanmaraş #Osmaniye #Adiyaman #Diyarbakır #sanlıurfa #Adana #Antakya #Ankar                 |
| İlgisiz | 0 | Deprem Suriye, Irak, İsrail, Filistin, Lübnan ve Kıbrıs'ta da hissedildi. Suriye ve Irak'ta da enkaz altında kalanlar var. #deprem #DEPREMOLDU #Syria #Iraq                                                                           |
| İlgisiz | 0 | Allah yardımcımız olsun. Lütfen telefon hatlarını ve anayolları gereksiz yere işgal etmeyelim.Yardıma ihtiyacı olanlara kolay ulaşılabilin. #Hatay #earthquake #Kandilli #sanlıurfa #Gaziantep #Malatya                               |
| İlgisiz | 0 | Yine kaybettğimiz canlar, yine depremle savaşıyoruz.. göçük altında kalanlara sağ sağlım ulaşmak ümidiyle.. geçmiş olsun Türkiye'm 🇹🇷 #Turkey #earthquake https://t.co/2ZH5I0s2                                                       |
| İlgisiz | 0 | #earthquake ömrümde ilk defa bu kadar korktum ya                                                                                                                                                                                      |
| İlgisiz | 0 | #Nurdağ merkezli #earthquake #deprem #Hatay, #Kahramanmaraş, #Adana, #Gaziantep, #Osmaniye, #Kilis, #sanlıurfa, #Diyarbakır, #Malatya, #Elazığ'daki binalarda yıkılanlar var, bu depremin hinterlandı çok büyük, Rabbimiz c           |
| İlgisiz | 0 | Geçmiş olsun Türkiye'm. Hindistan'dan iyi dilekler 11 🇮🇳 #earthquake                                                                                                                                                                  |

**Şekil 11.** İlgisiz olarak etiketli twitter mesajlarının ön işleme öncesi durumu

|         |   |                                                                                                                                                                                  |
|---------|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| İlgisiz | 0 | daha kötüsünü yaşamadım. kahramanmaraşta 7.4 büyüklüğünde bir deprem. enkaz altında olanlar var allahım acil yardım et o insanlara sen beterinden kuru. bir daha olmaz inş       |
| İlgisiz | 0 | arkadaşlar herkes adresler paylaşıyor rt ediyor bunlar doğru ama inanılmaz derecede yıkılan enkaz altında kalan bina var gibi oyuşten tüm illerden ekipler getirilmeli büyük çad |
| İlgisiz | 0 | kayseride depremin etkisi. çevre illerdeki etkisinin yayılması adına bir daha paylaşıyorum. cümleten geçmiş olsun.                                                               |
| İlgisiz | 0 | lütfen gerekmedikçe hatları meşgul etmeyin yardıma ihtiyacı olan insanlar var duyarlı olun biraz 7.8 şiddetinde deprem                                                           |
| İlgisiz | 0 | deprem suriye irak israil filistin lübnan ve kıbrıs'ta da hissedildi. suriye ve irakta da enkaz altında kalanlar var.                                                            |
| İlgisiz | 0 | allah yardımcımız olsun. lütfen telefon hatlarını ve anayolları gereksiz yere işgal etmeyelim.yardıma ihtiyacı olanlara kolay ulaşılabilin.                                      |
| İlgisiz | 0 | yine kaybettğimiz canlar yine depremle savaşıyoruz.. göçük altında kalanlara sağ sağlım ulaşmak ümidiyle.. geçmiş olsun türkiyem                                                 |
| İlgisiz | 0 | ömrümde ilk defa bu kadar korktum ya                                                                                                                                             |
| İlgisiz | 0 | merkezli daki binalarda yıkılanlar var bu depremin hinterlandı çok büyük rabbimiz devletimiz ve milletimizin yardımcısı olsun.allahım sen kuru                                   |
| İlgisiz | 0 | geçmiş olsun türkiyem. hindistandan iyi dilekler 11 🇮🇳                                                                                                                           |

**Şekil 12.** İlgisiz olarak etiketli twitter mesajlarının ön işleme sonrası durumu

Yapılan tüm bu işlemler sonra 3 farklı etiket ile etiketlenmiş 2000 verilik ve gürültüleri giderilmiş bir veri seti elde edilmiştir.

Yapılan çalışma kapsamında ilerleyen kısımlarda oluşturulan bu veri seti dışında internet ortamında yaygın olan 2 farklı veri seti daha kullanılmıştır. Bunlardan ilki tez kapsamında çalışılan konuya paralel içeriği olan CrisisNLP veri serinden oluşturulmuş bir alt veri setidir. Bu veri setinde dünya genelinde meydana gelen doğal afetler sonrası atılan Twitter mesajlarından oluşmaktadır. Bu veri setinden 20.000 verinin olduğu bir alt veri seti oluşturulmuştur ve Türkçe ye çevrilmiştir. Veri setindeki Twitter mesajları 2 etikete sahiptir. Bunlar “bilgilendirici” ve “bilgilendirici\_değil” şeklindedir. Bilgilendirici olanlar yaşanan doğal afet ile ilgili herhangi bir bilgi içeren Twitter mesajlarından oluşurken, bilgilendirici olmayanlar ise yaşanan afetten bağımsız atılan mesajlardan oluşmaktadır.

Kullanılan bir diğer veri seti ise tez kapsamında çalışılan konudan farklı içeriğe sahip Türkçe ürün yorumları içeren bir veri setidir. Bu veri setinde ise 7.584 adet veri bulunmaktadır ve internet ortamından hazır bir şekilde elde edilmiştir (Kaggle ürün yorumları veri seti). Veri seti 3 etiketten oluşmaktadır. Etiketleri nötr, olumlu ve olumsuz olmak üzere ayrılmıştır. Sonuç olarak tez kapsamında etiketleyerek oluşturduğumuz 6 Şubat Depremi veri seti, CrisisNLP’den alınarak oluşturulan diğer veri set ve Kaggle üzerinden alınan ürün yorumları veri seti olmak üzere 3 veri seti kullanılmıştır. Bu veri setlerine ait kelime çantaları Şekil 13, Şekil 14 ve Şekil 15’de verilmiştir.



**Şekil 13. 6** Şubat deprem verisi kelime bulutu



**Şekil 14.** Ürün yorumları içeren veri seti kelime bulutu



**Şekil 15.** CrisisNLP veri seti kelime bulutu

## Yöntem

Bu tez çalışmasında 6 Şubat 2023 Türkiye ve Suriye depremi konu alınarak olabilecek en hızlı şekilde az sayıda etiketli veri ile yüksek başarılı sonuçların alınması üzerine kurulmuştur. Bu amaç doğrultusunda üzerinde çalışılacak verinin kaynağı sosyal medya platformlarından aktif olarak kullanılan Twitter seçilmiştir. Twitter'ın seçilmesinde bazı avantaj ve dezavantajlar bulunmaktadır.

Twitter'ın seçilmesindeki avantajlar;

- Gerçek Zamanlı Veri: Twitter, gerçek zamanlı veri sağlar. Bu, depremler gibi acil durumlar hakkında anlık bilgi akışı elde etmek için idealdir.
- Büyük Veri Seti: Twitter'dan elde edilen veriler geniş ve çeşitlidir, bu da modelin daha kapsamlı ve genel bir öğrenme yapmasını sağlar.
- Toplumun Algısını Yansıtma: Twitter verileri, toplumun depremle ilgili görüşlerini ve tepkilerini doğrudan yansıtır, bu da sosyal dinamikler ve toplumun bilinç düzeyi hakkında bilgi sağlar.

Twitter'ın seçilmesindeki dezavantajlar;

- Gürültülü ve Düzensiz Veri: Twitter verileri genellikle gürültülü ve düzensizdir. Yanıltıcı bilgiler, argo ifadeler ve imla hataları veri temizliği ve ön işleme gereksinimini artırır.
- Sınırlı Bağlam: Twitter mesajları kısa ve bazen bağlamdan yoksundur, bu da metnin doğru bir şekilde sınıflandırılmasını zorlaştırır.
- Önyargı Riski: Twitter kullanıcıları belirli bir demografik veya sosyoekonomik gruba ait olabilir, bu da veri setinin temsil ediciliğini sınırlar.

Herhangi bir kriz anında veya hemen sonrasında geleneksel makine öğrenmesi yöntemleri ile çalışmak için yeterli sayıda veri toplamak oldukça zordur. Bu çalışmada yeterli çoğunlukta etiketli veri olmadığı durumlarda geleneksel yöntemlerin yetersiz kaldığı ve bu durumun önüne geçmek için transfer öğrenme yöntemlerinin daha iyi sonuçlar verebileceğini göstermek amaçlanmıştır. Geleneksel yöntemler ile transfer öğrenme yöntemlerinden olan BERT, RoBERTa ve DistilBERT modellerinin başarımları kıyaslanmıştır. Geleneksel yöntemler için Tf-Idf kelime gömme yöntemleri ile Lojistik Regresyon, Naive Bayes, Destek Vektör Makinesi (SVM), K-En Yakın Komşu (KNN) ve Karar Ağacı algoritmaları kullanılmıştır. Bunlara ek olarak eldeki etiketli verinin daha da az olması durumunda zero-shot ve few-shot uygulamalarının da sonuçları incelenmiştir. Bu inceleme için de BERT ve TARS modelleri kullanılmıştır.

### **Geleneksel makine öğrenmesi yöntemlerinin uygulanması**

Geleneksel makine öğrenmesi yöntemleri, yapay zekâ ve veri bilimi alanlarında temel bir taş oluşturmaktadır. Bu yöntemler, belirli bir problemi çözmek için veri tabanlı kararlar almak üzere tasarlanmış algoritmaları içermektedir. Geleneksel makine öğrenimi, derin öğrenme ve nöral ağlar gibi daha yeni tekniklerin gelişinden önce ortaya çıkmıştır ve genellikle daha basit, ama etkili modeller sunmaktadır. Çalışma kapsamında az sayıda etiketli verinin

olması durumunda da etkili sonuçlar verip veremeyeceği incelenmiştir. Bu inceleme sonucunda transfer öğrenme yöntemleri ile kıyaslanmaktadır.

Geleneksel yöntemlerin uygulanması için Lojistik Regresyon, Naive Bayes, Destek Vektör Makinesi (SVM), K-En Yakın Komşu (KNN) ve Karar Ağacı olacak şekilde 5 algoritma seçilmiştir. Bu algoritmaların seçilmesinde bazı faktörler etkili olmuştur. Bunlardan başlıca olanları; Destek Vektör Makinesi algoritması, karmaşık sınıflandırma problemlerinde güçlü performans göstermektedir. Yüksek boyutlu veri setlerini işleyebilir ve marjinal ayrımı maksimize ederek sınıflandırma yapmaktadır. Benzer şekilde K-En Yakın Komşu algoritması, basit ve anlaşılması kolay bir algoritmadır. Veri etiketlemede ve regresyon problemlerinde iyi sonuçlar vermektedir. Parametre ayarlaması ve veri ölçeklendirmesi ile esnek bir şekilde kullanılmaktadır. Karar Ağacı algoritması hem sınıflandırma hem de regresyon problemleri için kullanılmaktadır. Ayrıca, veri özelliklerinin önemini anlamada yardımcı olmaktadır.

Çalışmada incelenen metin verileri, makinenin işleyebilmesi için özellik çıkarımı sürecinden geçirilmiştir. Metnin içeriğini temsil etmek ve makinenin anlayabileceği bir formata dönüştürmek için Tf-Idf kelime gömme tekniği kullanılmıştır. Bu teknik, metin verisinin analizinde ve modellemede önemli rol oynamaktadır. Bu algoritmaların uygulanmasında öncelikle hazırlanan 2000 verilik veri setinin tamamı %80'i eğitim %20'si test olacak şekilde dengeli olarak bölünmüş ve eğitimler yapılmıştır. Ardından daha az etiketli veri ile nasıl bir sonuç vereceğini gözlemlemek içinde eğitim verisindeki sayı dengeli ama rasgele seçilmiş 100 veri olacak şekilde değiştirilmiş ve bu şekilde de eğitimler yapılarak sonuçlar alınmıştır. Bu uygulamaların yanı sıra, modelin genel performansının doğru bir şekilde değerlendirilmesi amacıyla k-fold çapraz doğrulama (cross-validation) tekniği kullanılmıştır. Bu süreçte, veri seti rastgele beş eşit parçaya bölünmüş (k değeri 5 olarak belirlenmiştir) ve her iterasyonda dört parça eğitim verisi, kalan parça ise test verisi olarak kullanılmıştır. Bu yöntem sayesinde, her bir parçanın bir kez test verisi olarak değerlendirilmesi sağlanarak, modelin farklı veri alt kümelerindeki performansı etkin bir şekilde incelenmiştir.

### **Transfer öğrenme yöntemlerinin uygulanması**

Transfer öğrenme alanında ön plana çıkan üç temel model olan BERT, RoBERTa ve DistilBERT incelenmiştir. Bu modeller, geniş ve çeşitlendirilmiş veri setleri üzerinde önceden eğitilmişlerdir, bu sayede dilin karmaşık yapısını ve nüanslarını anlamak için geniş bir temel sağlarlar. BERT, çift yönlü bir yapıya sahiptir ve metinlerin daha kapsamlı bir şekilde analiz edilmesini mümkün kılar. RoBERTa ise BERT'in geliştirilmiş bir sürümüdür ve daha geniş veri setleri ve uzun eğitim süreçleri kullanarak dil modellemesindeki etkinliği artırır. Öte yandan, DistilBERT, BERT'in sadeleştirilmiş bir versiyonudur ve hafifliği sayesinde daha az hesaplama

kaynağı gerektiren durumlar için uygundur. Bu modeller, özgün eğitimlerinin ardından spesifik NLP görevleri için ince ayar (fine-tuning) yapılabilirler, böylece belirli bir göreve veya veri setine özel ihtiyaçlara uygun hale getirilir. İnce ayar süreci, modellerin önceden öğrenmiş oldukları genel dil yapısını, özel görevin gereksinimleriyle bütünleştirerek daha yüksek doğruluk ve verimlilik sağlamaktadır.

### ***BERT, RoBERTa ve DistilBERT'in uygulanması***

Bu çalışmada metin sınıflandırma problemi için önceden eğitilmiş BERT modelinin Türkçe versiyonu tercih edilmiştir. Bu model, oluşturulan veri seti ile ek eğitim (fine-tuning) yapılarak göreve özelleştirilmiştir. Toplam veri seti, rastgele olarak %80 eğitim ve %20 test olacak şekilde iki alt küme olarak bölünmüştür. “dbmdz/bert-base-turkish-uncased” adlı, Türkçe için önceden eğitilmiş BERT modeli kullanılarak metinler tokenlara dönüştürülmüştür. Metinler tokenizasyon işlemine tabi tutulmuş, bu sırada maksimum token sayısı olarak modelin default sınırı benimsenmiştir. Bu sınıra uymayan metinler kesilirken, daha kısa metinler padding ile bu uzunluğa getirilmiştir. Modelin eğitimi sırasında kullanılan hiperparametreleri ayarlanmıştır. “per\_device\_train\_batch\_size” değeri 8 olarak belirlenmiştir. Bu seçim hem hızlı eğitim süreci sağlamak hem de hafıza sınırlamalarını aşmamak için yapılmıştır. “per\_device\_eval\_batch\_size” parametresi her GPU veya CPU için değerlendirme sırasında kullanılacak örnek sayısıdır ve 8 olarak belirlenmiştir. Böylece eşit bir eğitim-değerlendirme dengesi sağlanmıştır. 20 çevrim kullanılarak, modelin veriyi yeterince öğrenmesi ama aynı zamanda aşırı öğrenmeye maruz kalmaması amaçlanmıştır. RoBERTa ve DistilBERT içinde benzer parametreler kullanılırken farklı olarak RoBERTa için “roberta-base”, DistilBERT için ise “distilbert-base-uncased” adlı önceden eğitilmiş modeller kullanılmıştır.

### **Sıfır atış öğrenme (Zero-Shot) - birkaç örnek öğrenme (Few-Shot) yöntemlerinin uygulanması**

Az sayıda etiketli verinin olması durumunda yüksek başarımlı sonuçlar elde etmek söz konusu olduğunda sıfır atış ve birkaç örnek öğrenme kavramlarının uygulanması ihtiyacı duyulmuştur. Çalışma kapsamında 2000 verilik bir etiketli veri seti oluşturulmuştur. Fakat bir kriz anında bu sayıda etiketli bir veri seti oluşturmak yeteri kadar kısa sürmeyebilir. Bu nedenle hiç etiketli verimiz yokken önceden eğitilmiş bir modelde başarılı bir sınıflandırma yapılabilir mi sorusunun cevabı için sıfır atış yöntemi uygulanmıştır. Bunun yanında birkaç örnek ile yapılan eğitimlerin sonuçları da incelenmiştir. Örnek olarak K değeri 1,5,10,15,50 ve 100 olarak seçilerek kademeli bir şekilde artırılmıştır. Bu uygulamalar için 2 model seçilmiştir. Bunlardan ilki BERT ve ikincisi ise TARS modelidir. BERT modelinde hem sıfır atış hem de

birkaç öğrenme uygulanırken; TARS ile sadece birkaç atış öğrenme uygulanmıştır. Her iki modelde de çapraz doğrulama (cross validation) yapılmıştır.

### ***BERT'in uygulanması***

BERT modelinin seçilmesinin ardından yapılan ilk adım sıfır atış öğrenme olmuştur. BERT ön eğitilmiş bir model olduğu için herhangi bir eğitim veya ince ayar yapılmamıştır. Geniş kapsamlı Türkçe veriler ile eğitilmiş model olan “dbmdz/bert-base-turkish-uncased” modeli seçilmiştir. Ardından 2000 adet verinin tamamı sınıflandırılmak üzere modele verilmiştir. 3 farklı etiket ile sınıflandırılarak sonuçlar not edilmiştir. Ardından birkaç örnek ile öğrenme yöntemi için yine aynı ön eğitilmiş “dbmdz/bert-base-turkish-uncased” modeli alınmıştır. Fakat bu sefer farklı olarak  $K = 1, 5, 10, 15, 50, 100$  olacak şekilde her sınıftan belirli sayıda örnekler verilmiştir. Her  $K$  değeri için 20 çevrim ile eğitimler yapılmış ve sonuçlar not edilmiştir. Son olarak CrisisNLP (crisis\_datasets\_benchmarks) veri setinden alınan 20.000 doğal afet verisi elde edilmiş ve bu veri seti Türkçe diline çevrilmiştir. Bu veri seti ile BERT'in ön eğitilmiş modeline bir ince ayar (fine-tune) işlemi uygulanarak ikinci bir eğitim yapılmıştır. Ardından çalışma kapsamında oluşturulan veri setinden  $K = 1, 5, 10, 15, 50, 100$  olacak şekilde her etiketten örnekler verilerek ikinci bir ince ayar (fine-tune) işlemi yapılmıştır. Farklı veri setinin de deneylere dahil edildiği çalışma tekrarlanmış, bu sefer benzer konuda bilgiler içeren veri seti yerine konudan bağımsız içeriğe sahip bir veri seti kullanılmıştır. Bu veri seti ürün yorumlarından oluşan bir veri setidir. Burada amaç benzer bilgiler içeren bir veri seti ile ön eğitilmiş modelde ince ayar (fine-tune) yapılmasının veya konudan bağımsız bilgiler içeren bir veri setini ile ince ayar (fine-tune) yapılmasının birkaç örnek ile öğrenme sonuçlarına etkisinin ölçülmesidir. Tüm bu deneyler sonrası alınan sonuçlar için başarı kıyaslaması yapılmıştır.

### ***TARS'ın uygulanması***

TARS, Flair adında, açık kaynak kodlu bir doğal dil işleme çerçevesi kullanarak, metin sınıflandırma ve dil modelleme görevlerine odaklanmış bir şekilde tasarlanmıştır (Akbik *et al.* 2019). Flair, doğal dil işleme (NLP) alanında eğitim ve model dağıtımını kolaylaştırmak için tasarlanmış bir çerçevedir. Bu çerçeve, kelime ve doküman gömülmesi (embedding) gibi konsept olarak farklı tipte verilere basit ve birleşik bir arayüz sunmaktadır (Akbik *et al.* 2019). Bu çalışmada birkaç örnek öğrenme yöntemi ile TARS modeli kullanılmıştır ve oluşturulan veri setinde sınıflandırma işlemi yapılmıştır. TARS eğitim sırasında, her metin örneği için <tasklabel, text> çifti oluşturur. Doğru etiket için True, yanlış etiketler için False olacak şekilde atamalar yapar. Çok sınıflı problemlerde, True olarak tahmin edilen en yüksek güven düzeyine

sahip sınıf nihai tahmin olarak seçilmektedir. Bir duygu sınıflandırma probleminde iki tür eğitim çifti oluşturulur. Örneğin, "Bu film harikaydı" cümlesi için:

<"positivesentiment", "Bu film harikaydı"> → TRUE

<"negativesentiment", "Bu film harikaydı"> → FALSE

çiftleri oluşturulur. Her çift için, etiket ve metin çapraz dikkat mekanizmasıyla birleştirilir ve ikili bir sınıflandırma yapılır. Bir metin sınıflandırılacağı zaman, model her olası etiket için bir tahmin yapar. Yani, her etiket için metin ile değerlendirilir:

<"positivesentiment", "Yeni film çok kötüydü">

<"negativesentiment", "Yeni film çok kötüydü">

Model, bu iki durum için True veya False döndürür ve en yüksek olasılığa sahip etiket seçilir. TARS modeli sınıflandırma problemlerinde metin ve etiketi birleştirme aşamasında BERT algoritmasının çapraz dikkat mekanizmasını kullanmaktadır (Halder *et al.* 2020).

Kaynak görev olarak, BERT'te olduğu gibi CrisisNLP (crisis\_datasets\_benchmarks) veri setinden alınan 20.000 doğal afet verisi Türkçe diline çevrilerek kullanılmıştır. Aynı zamanda konu benzerliğinin birkaç örnek öğrenmeye etkisini bu uygulamada da kıyaslamak için ürün yorumlarından oluşan veri seti kullanılarak deney tekrar edilmiştir. Bu çalışmada TARS modeli ile birkaç atış öğrenme uygulamasında hedef görev olarak çalışma kapsamında oluşturulan yeni veri seti olmuştur (Kaggle veri setinden alınarak oluşturulan Türkçe veri seti.). Sınıflandırma görevlerinde her sınıftan K tane örnek içeren eğitim seti oluşturulmuştur ve kalan veriler modeli değerlendirmek üzere kullanılmıştır.

## ARAŞTIRMA BULGULARI VE TARTIŞMA

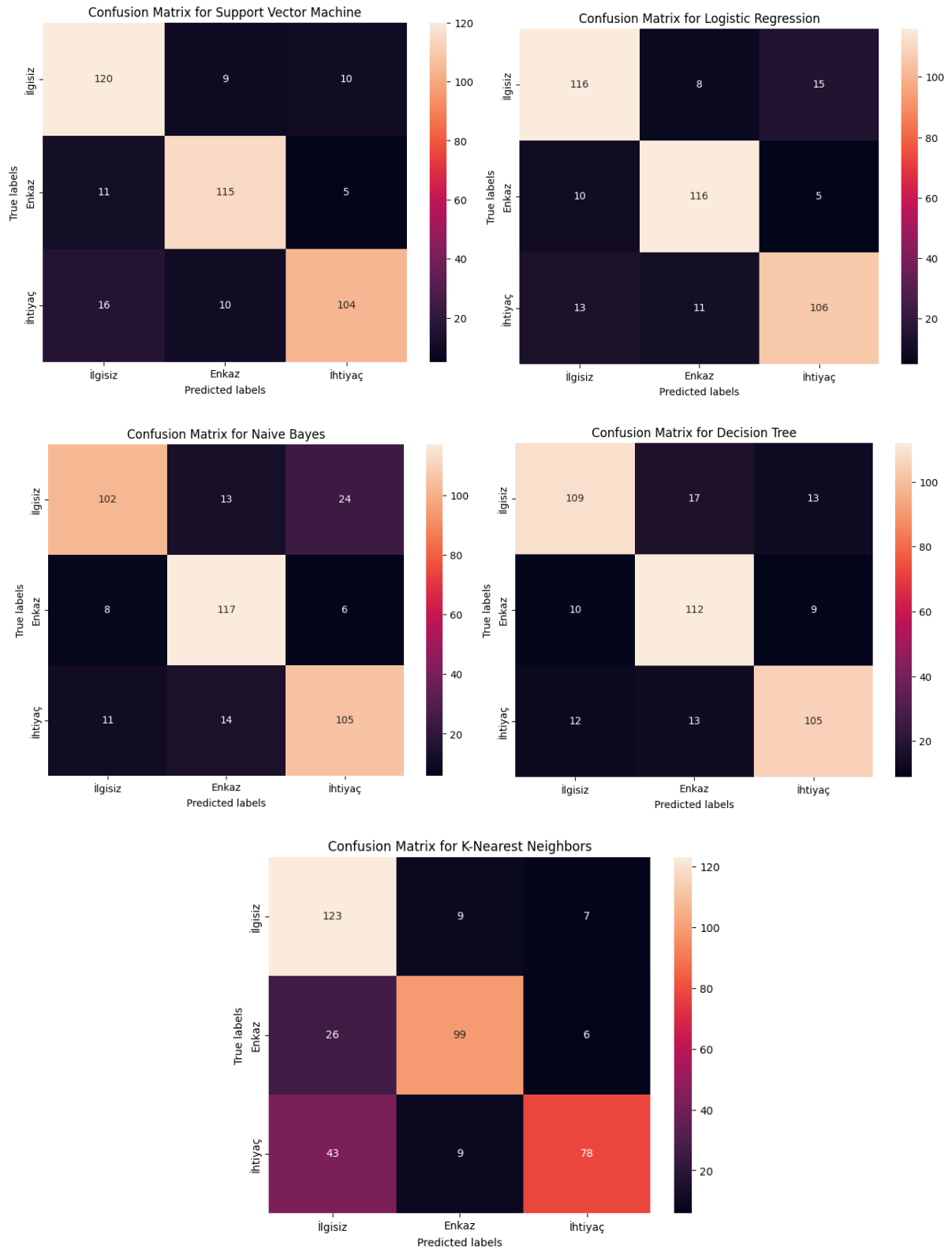
Çalışma kapsamında amaçlanan hedef doğrultusunda ön eğitilmiş modeller kullanılıp az sayıda etiketli veri ile ince ayar sonucu eğitim ve test sonuçları alınmıştır. Bir deprem sonrasında sayıca fazla veri toplamak ve bunlara ihtiyaca yönelik etiketler vererek sıfırdan eğitimler yapmak oldukça maliyetli olacağı gibi aksiyon alınması için yeterli süre olmayacaktır. Bu durumun önüne geçmek ve deprem sonrasında gelen bilgileri başarılı bir şekilde sınıflandırıp daha sonra bu sınıfları farklı amaçlar doğrultusunda kullanmak için transfer öğrenme ile sıfır-atış / birkaç örnek öğrenme yöntemleri oldukça etkilidir.

### Geleneksel Yöntemler ile Sınıflandırma Sonuçları

Çalışmanın bu kısımda daha önce oluşturulan 2000 adet Türkçe veriden oluşan veri seti kullanılarak deneyler yapılmıştır. Tf-Idf kelime gömme yöntemi ile geleneksel yöntemlerden KNN, Naive Bayes, Lojistik Regresyon, Karar Ağacı ve SVM algoritmaları kullanılmıştır. TfidfVectorizer kullanılarak TF-IDF vektörleri oluşturulmuştur. max\_features=5000 parametresi ile oluşturulan TF-IDF vektörlerinin her birinde en fazla 5000 farklı kelimeyi içermesi sağlanmıştır. Bu, TF-IDF matrisinin her bir satırının (her bir belge için) 5000 özellik (kelime) içereceği ve bu özelliklerin her birinin belirli bir kelimenin belgedeki önemini yansıtacağı anlamına gelir. Eğitimler yapılırken ilk önce 2000 veri her etiketten eşit dağılımda olacak şekilde kullanılmıştır. Ardından transfer öğrenme yöntemleri ile kıyaslanabilir olması için eğitim için etiket dağılımının dengeli olduğu 100 veri kullanılmıştır. Test için veri setinin geri kalanı (1900 veri) kullanılmıştır.

**Tablo 6.** Geleneksel Yöntemler ile Sınıflandırma Sonuçları

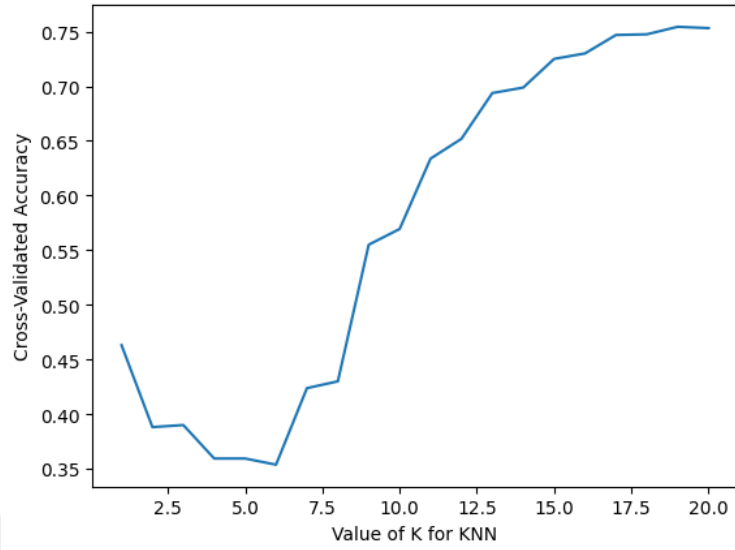
|                                        | Model              | Doğruluk       | Veri Seti       | Veri boyutu                                                     |
|----------------------------------------|--------------------|----------------|-----------------|-----------------------------------------------------------------|
| <b>Geleneksel Yöntemler (Tf - Idf)</b> | SVM                | %82,94         | 6 Şubat Depremi | 2000 veri<br>Çapraz Doğrulama<br>(Cross-Validation)<br>(5-Fold) |
|                                        | KNN (K=19)         | %75.87         | 6 Şubat Depremi |                                                                 |
|                                        | Lojistik Regresyon | % <b>82,99</b> | 6 Şubat Depremi |                                                                 |
|                                        | Naive Bayes        | %80,00         | 6 Şubat Depremi |                                                                 |
|                                        | Karar Ağacı        | %77.31         | 6 Şubat Depremi |                                                                 |



**Şekil 16.** Geleneksel yöntemlere ait karmaşıklık matrisi (Yüksek eğitim verisi)

Matrisler incelediğinde, Naive Bayes algoritmasının diğer algoritmalara göre ilgisiz ve ihtiyaç kategorilerini ayırt etmekte biraz zorlandığı görülmektedir. Diğer yandan KNN algoritmasında ihtiyaç sınıfı en çok yanlış tahmin edilen sınıf olduğu görülmektedir: 78 doğru tahmin varken, 43'ü ilgisiz ve 9'u enkaz olarak yanlış sınıflandırılmış. Burada ihtiyaç ve ilgisiz sınıfının ayırımında zorlandığı görülmektedir. KNN algoritmasında en iyi sonuç alınan K

değerinin belirlenmesi işlemi uygulanmıştır. K değerine ait optimal değer grafiği Şekil 17’te gösterilmektedir.

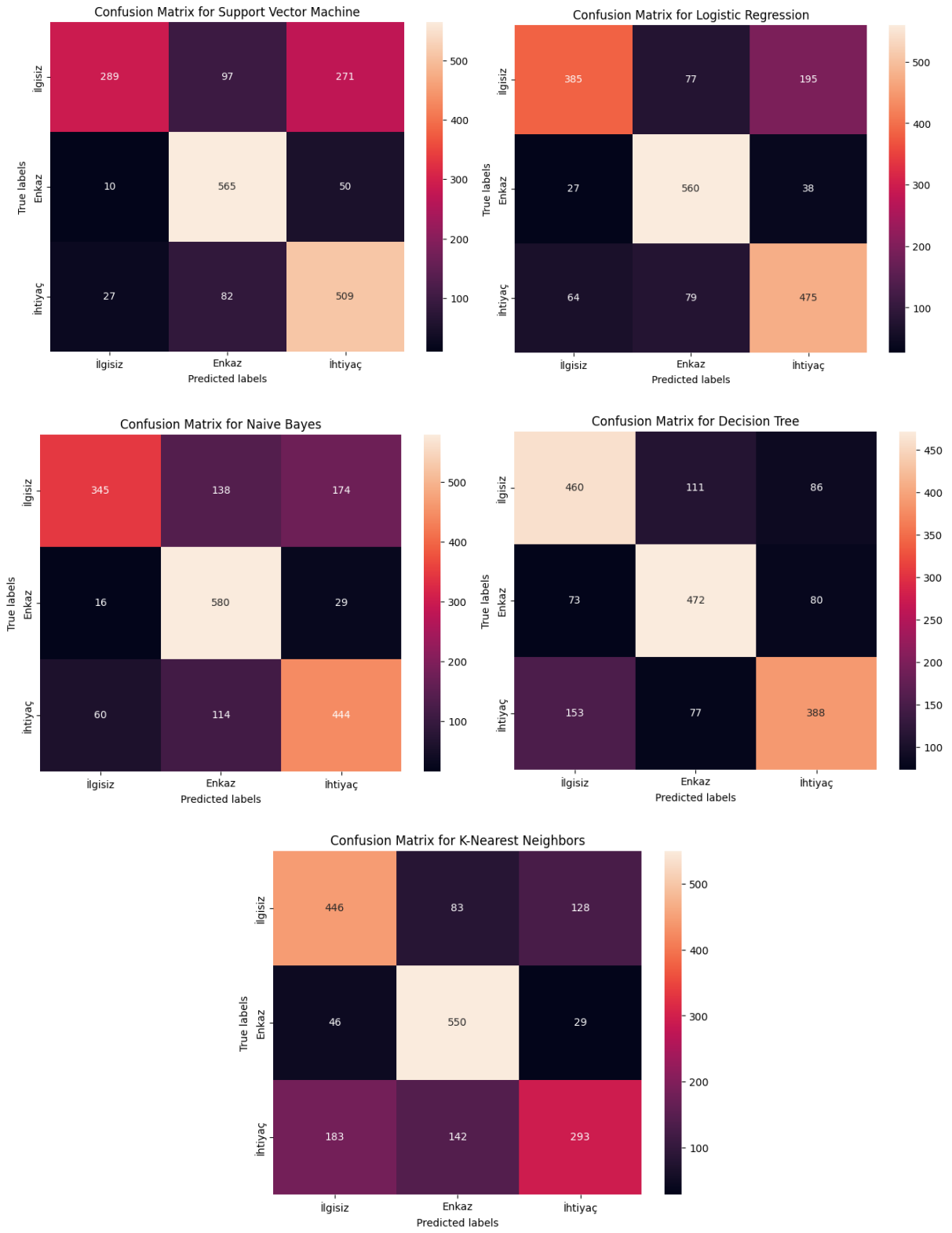


Şekil 17. Optimal K değeri için grafik (K=19)

Tablo 7. Geleneksel Yöntemler ile Sınıflandırma Sonuçları (100 Eğitim Verisi)

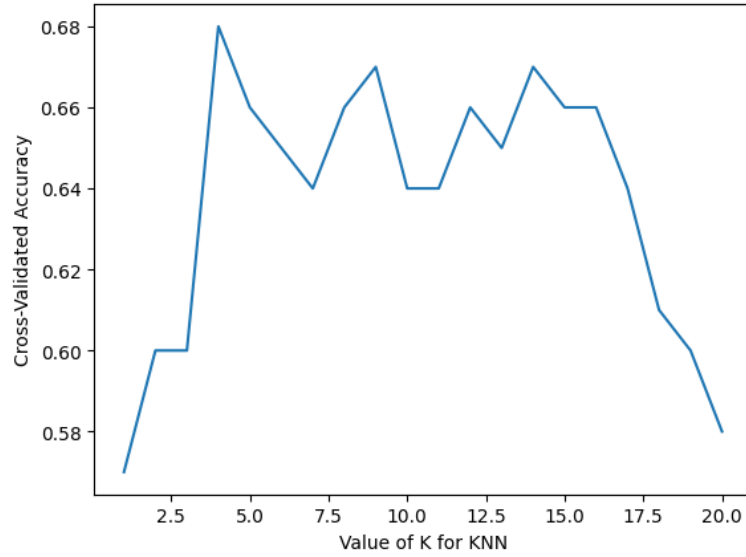
|                                        | Model              | Doğruluk | Veri Seti       | Veri Boyutu                           |
|----------------------------------------|--------------------|----------|-----------------|---------------------------------------|
| <b>Geleneksel Yöntemler (Tf - Idf)</b> | SVM                | %71,74   | 6 Şubat Depremi | Eğitim (100 veri)<br>Test (1900 veri) |
|                                        | KNN (K=4)          | %67,84   | 6 Şubat Depremi |                                       |
|                                        | Lojistik Regresyon | %74,73   | 6 Şubat Depremi |                                       |
|                                        | Naive Bayes        | %72.05   | 6 Şubat Depremi |                                       |
|                                        | Karar Ağacı        | %69,47   | 6 Şubat Depremi |                                       |

Eğitim verisinin boyutu düşürülerek eğitim yapıldığında alınan sonuçlar Tablo 7’de verilmiştir. Sonuçlar incelendiğinde eğitim verisinin boyutundaki düşüş sonuçlarda da gözle görülür bir düşüşe neden olmuştur. Az veri ile eğitimde geleneksel yöntemlerde en iyi sonucu Lojistik Regresyon algoritması vermiştir. Sonuçlara ait karmaşıklık matrisleri Şekil 18’te verilmiştir.



**Şekil 18.** Geleneksel yöntemlere ait karmaşıklık matrisi (Düşük eğitim verisi)

KNN algoritmasında eğitim verisinin boyutu küçültüldüğünde de en iyi sonuç alınan K değerinin belirlenmesi işlemi uygulanmıştır. K değerine ait optimal değer grafiği Şekil 19’te gösterilmektedir.



**Şekil 19.** Optimal K değeri için grafik (K=4)

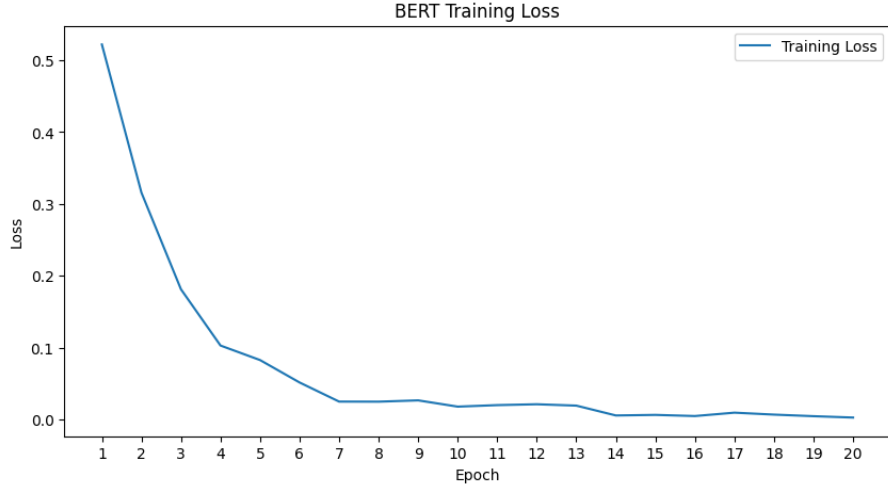
### Transfer Öğrenme Yöntemleri ile Sınıflandırma Sonuçları

Bu kısımda metin sınıflandırma problemi için önceden eğitilmiş BERT modelinin Türkçe versiyonu tercih edilmiştir. Bu model, oluşturulan veri seti ile ek eğitim (fine-tuning) yapılarak göreve özelleştirilmiştir. Toplam veri seti, rastgele ama eşit etiket dağılımı olacak şekilde %80'i eğitim ve %20'si test olarak bölünmüştür. Çevrim sayısı 20 seçilerek eğitim yapılmış ve en iyi sonuç veren kaydedilmiştir. Aynı zamanda her bir çevrim için kaybı görmek amacıyla loss grafiği elde edilmiştir. Sonuçlar Tablo 8'de verilmiştir.

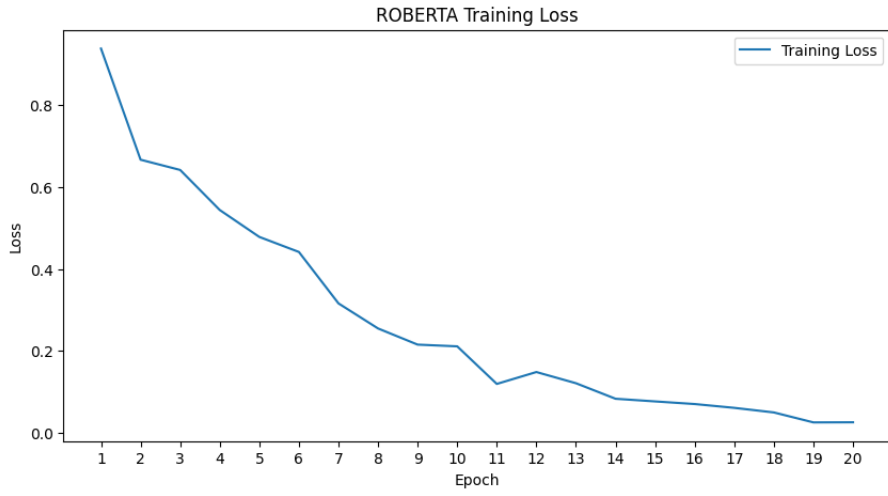
**Tablo 8.** BERT, RoBERTa ve DistilBERT Sonuçları

|                         | Model      | Doğruluk | Veri Seti       | Veri Boyutu            |
|-------------------------|------------|----------|-----------------|------------------------|
| <b>Transfer Öğrenme</b> | BERT       | %87,25   | 6 Şubat Depremi | %80 Eğitim             |
|                         | RoBERTa    | %84,25   | 6 Şubat Depremi | (1600 veri)            |
|                         | DistilBERT | %84,75   | 6 Şubat Depremi | %20 Test<br>(400 veri) |

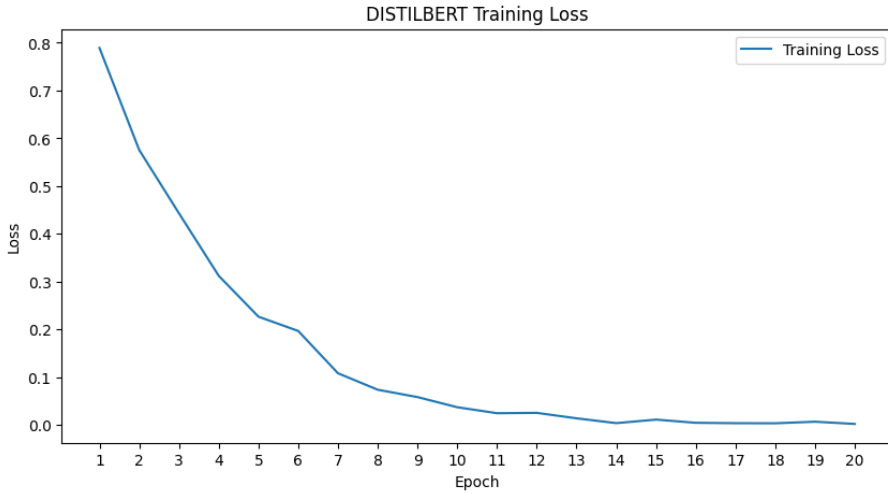
Sonuçlar incelendiğinde en iyi sonucu BERT'in verdiği görülmüştür. 20 çevrim için her bir modelin loss grafiği elde edilmiştir ve grafikler sırasıyla Şekil 20, Şekil 21 ve Şekil 22'de gösterilmektedir.



Şekil 20. BERT eğitim loss grafiği



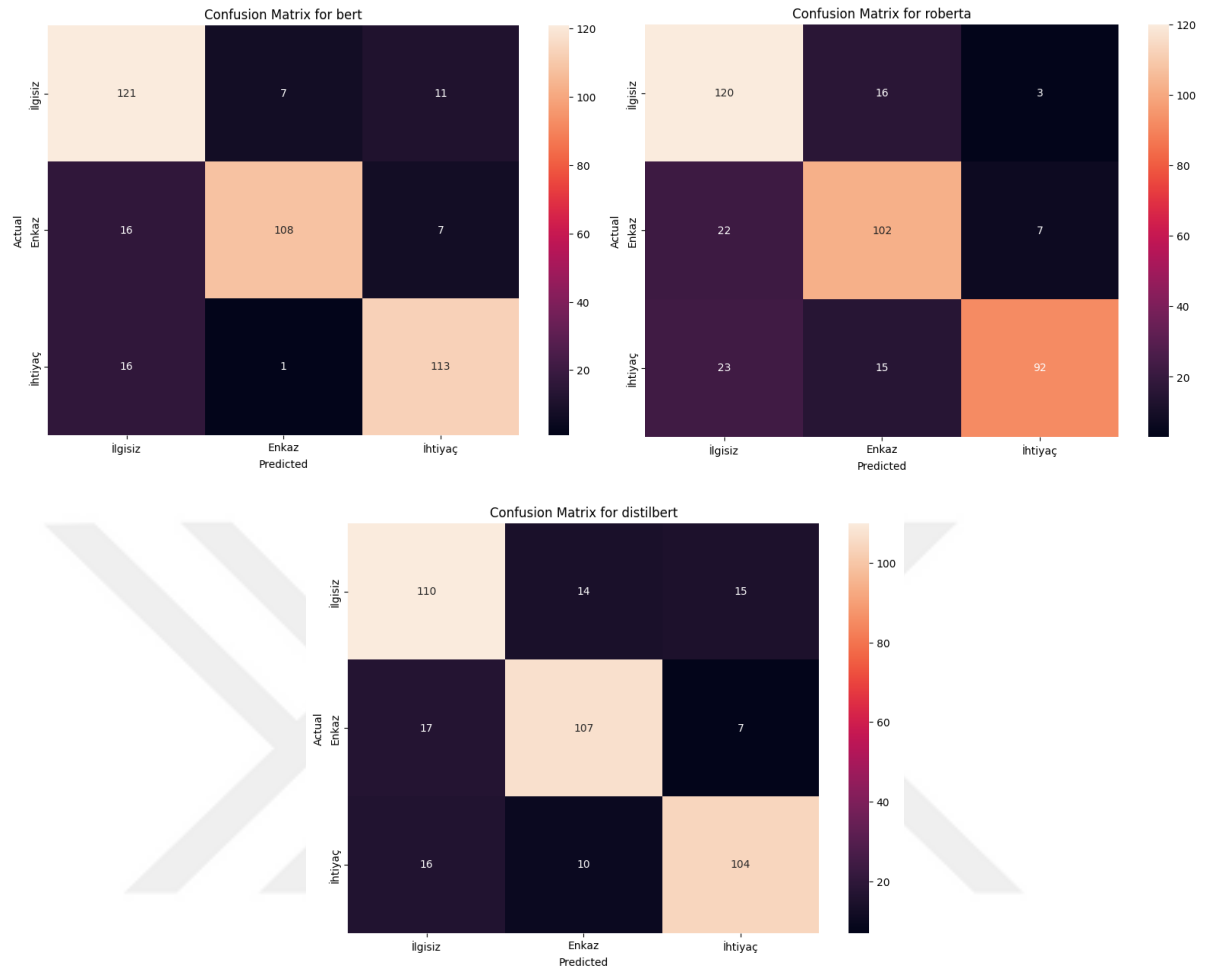
Şekil 21. RoBERTa eğitim loss grafiği



Şekil 22. DistilBERT eğitim loss grafiği

Elde edilen grafikler incelendiğinde tüm modellerde eğitim kayıplarının zamanla azaldığı ve belirli bir noktadan sonra kayıp değerlerinin stabil hale geldiği görülmektedir. Bu, modellerin eğitim süresince öğrenme kapasitelerinin iyileştiğini ve belirli bir çevrim sayısından

sonra ek eğitimin modelin performansını önemli ölçüde artırmadığını göstermektedir. Modellere ait karmaşıklık matrisleri Şekil 20’de verilmiştir.



**Şekil 23.** BERT, RoBERTa ve DistilBERT için karmaşıklık matrisleri

Bu üç karışıklık matrisi, BERT, RoBERTa ve DistilBERT modellerinin sınıflandırma performanslarını göstermektedir. BERT genel olarak iyi performans göstermiştir. Özellikle "İlgisiz" ve "İhtiyaç" sınıflarında yüksek doğruluk sağlamıştır. RoBERTa, "İlgisiz" sınıfında en iyi performansı gösterirken, diğer sınıflarda biraz daha fazla karışıklık yaşanmıştır. DistilBERT, üç sınıfın her birinde dengeli bir doğruluk sergilemiş, ancak "İlgisiz" ve "İhtiyaç" arasında bazı karışıklıklar olmuştur.

Geleneksel yöntemlerde yapıldığı gibi bu kısımda da çapraz doğrulama (cross validation) yapılmış ve sonuçlar incelenmiştir. Sonuçlar Tablo 9’da verilmiştir.

**Tablo 9.** BERT, RoBERTa ve DistilBERT İçin Çapraz Doğrulama Sonuçları

|                         | Model      | Doğruluk | Veri Seti       | Yöntem                                             |
|-------------------------|------------|----------|-----------------|----------------------------------------------------|
| <b>Transfer Öğrenme</b> | BERT       | %83,30   | 6 Şubat Depremi | Çapraz Doğrulama<br>(Cross-Validation)<br>(5-Fold) |
|                         | RoBERTa    | %79,46   | 6 Şubat Depremi |                                                    |
|                         | DistilBERT | %83.05   | 6 Şubat Depremi |                                                    |

**Sıfır Atış ve Birkaç Örnek Öğrenme ile Sınıflandırma Sonuçları**

Sıfır atış ve birkaç örnek ile öğrenme yöntemleri yeterli sayıda etiketli verinin bulunmadığı durumlar için geliştirilmiştir. Deprem anı ve hemen sonrası düşünüldüğünde yeterli sayıda etiketli veri olmama ihtimali oldukça yüksektir. Bu yöntemlerin bahsedilen durumlarda başarısının nasıl olduğu incelenmiştir. Bu kapsamda BERT(ft) ve TARS modelleri seçilmiştir.

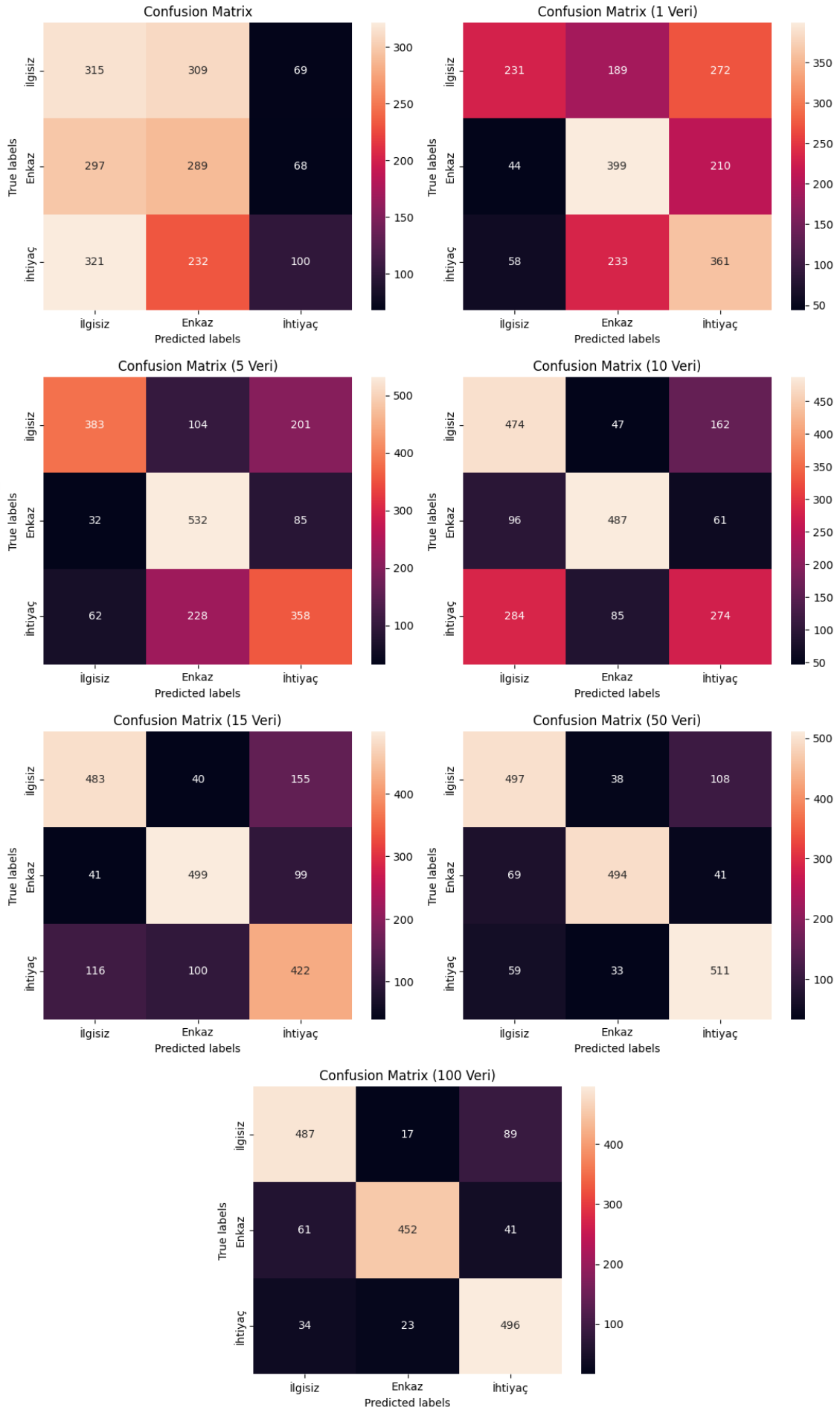
**BERT(ft) için sınıflandırma sonuçları**

Bu kısımda öncelikle BERT'in ön eğitilmiş modeli ele alınmıştır (dbmdz/bert-base-turkish-uncased). Bu model üzerinde herhangi bir ince ayar (fine-tune) veya eğitim yapılmadan ön eğitilmiş modelde çalışma kapsamında oluşturulan 3 etiketli 2000 verilik Türkçe veri seti sınıflandırılarak sıfır atış yöntemi uygulanmıştır. Sıfır atış yöntemi ile birkaç örnek ile öğrenme yönteminin karşılaştırılması açısından modele K değeri 1,5,10,15,50 ve 100 olacak şekilde örnekler verilerek eğitim ve sınıflandırma işlemi yapılmıştır. Sonuç olarak sıfır atış ile alınan sonuca göre birkaç örnek ile öğrenme daha iyi sonuçlar vermiştir. Bu sonuçlar Tablo 10'de verilmiştir.

**Tablo 10.** BERT(ft) Sıfır Atış ve Birkaç Örnek ile Öğrenme Sonuçları

|                  | K   | Doğruluk | Veri Seti       |
|------------------|-----|----------|-----------------|
| <b>BERT (ft)</b> | 0   | %35,20   | -               |
|                  | 1   | %49.62   | 6 Şubat Depremi |
|                  | 5   | %64,13   | 6 Şubat Depremi |
|                  | 10  | %63,51   | 6 Şubat Depremi |
|                  | 15  | %71,81   | 6 Şubat Depremi |
|                  | 50  | %81,18   | 6 Şubat Depremi |
|                  | 100 | %84.41   | 6 Şubat Depremi |

Bu sonuçlara ait karmaşıklık matrisleri aşağıda verilmiştir.



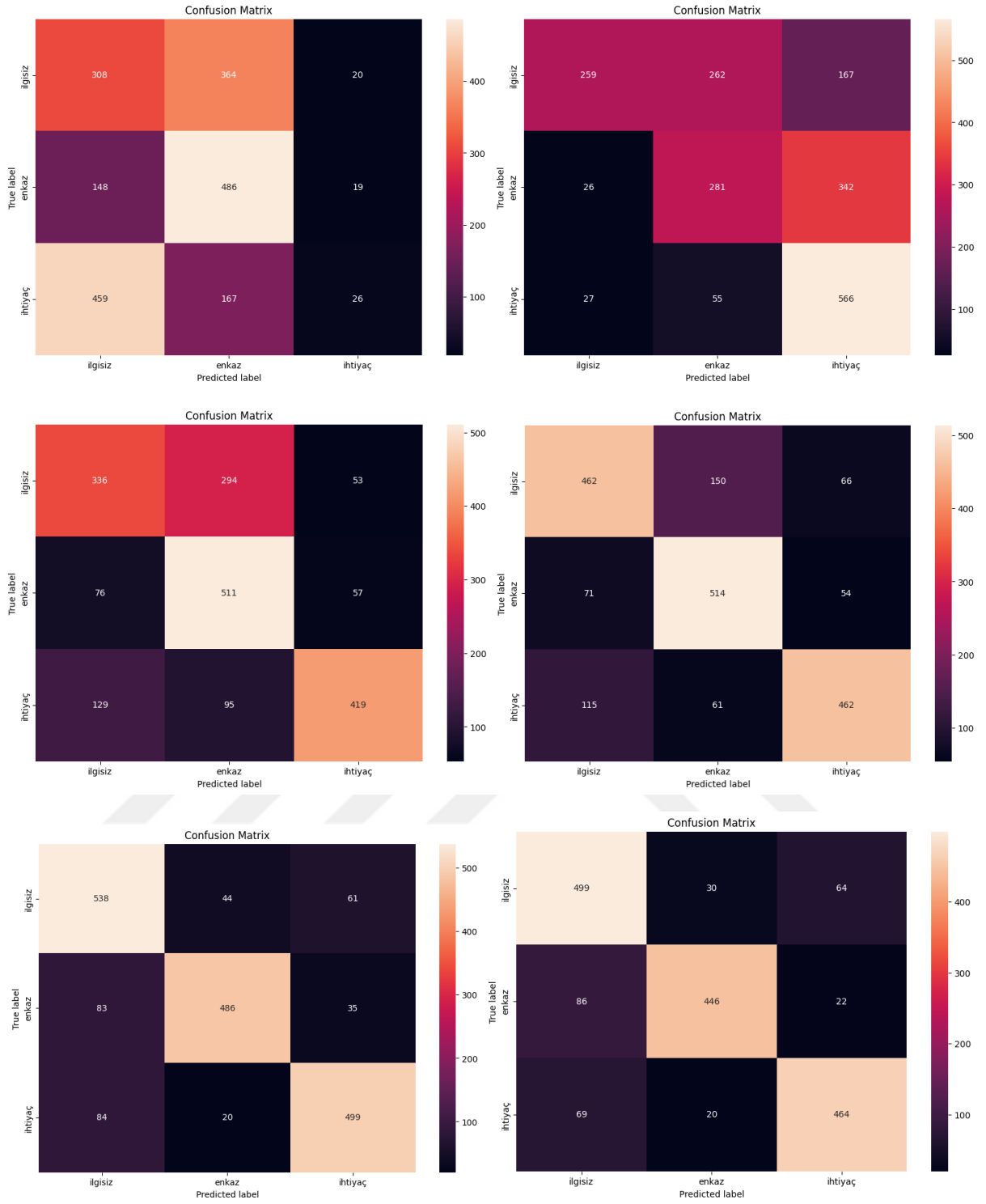
Şekil 24. BERT(ft) sıfır atış ve birkaç örnek öğrenme karmaşıklık matrisi

Ardından daha önce hazırlanmış internet ortamında hazır olan ve konu alınan probleme benzer konu içeren veri seti ile eğitim yaparak ince ayar yapmanın sonuçlara etkisi incelenmiştir. Bu doğrultuda ön eğitilmiş model CrisisNLP (crisis\_datasets\_benchmarks) veri setinden alınan 20.000 doğal afet verisi elde edilmiş Türkçe veri seti kullanılarak eğitilmiş ve ilk ince ayar işlemi yapılmıştır (Bu veri seti 2 etiketten oluşmaktadır.). Bu eğitim sonrasında birkaç örnek öğrenme yöntemi uygulanarak K değeri 1,5,10,15,50 ve 100 olacak şekilde tekrar eğitimler yapılmış ve sonuçlar incelenmiştir. Sonuçlar Tablo 11’de verilmiştir.

**Tablo 11.** BERT(ft) Farklı Veri Seti ile Eğitim Sonrası Birkaç Örnek Öğrenme Yöntemi Sonuçları

|                  | K   | Doğruluk      | Veri Seti                |
|------------------|-----|---------------|--------------------------|
| <b>BERT (ft)</b> | 1   | %41,06        | CrisisNLP + 6 Şubat (ft) |
|                  | 5   | %55,71        | CrisisNLP + 6 Şubat (ft) |
|                  | 10  | %64,26        | CrisisNLP + 6 Şubat (ft) |
|                  | 15  | %73,55        | CrisisNLP + 6 Şubat (ft) |
|                  | 50  | %82,32        | CrisisNLP + 6 Şubat (ft) |
|                  | 100 | <b>%82,89</b> | CrisisNLP + 6 Şubat (ft) |

Alınan sonuçlara göre farklı bir veri seti ile eğitim yaparak ardından birkaç örnek öğrenme ile ince ayar yapmanın olumlu bir etkisi olmamıştır. Bunun nedeni CrisisNLP’den elde edilen veri setinin 2 etiketten oluşurken; çalışma kapsamında oluşturulan ve ele alınan 6 Şubat Deprem veri setinin 3 etiketten oluşmasıdır. Bu sonuçlara ait karmaşıklık matrisi Şekil 25’te verilmiştir.



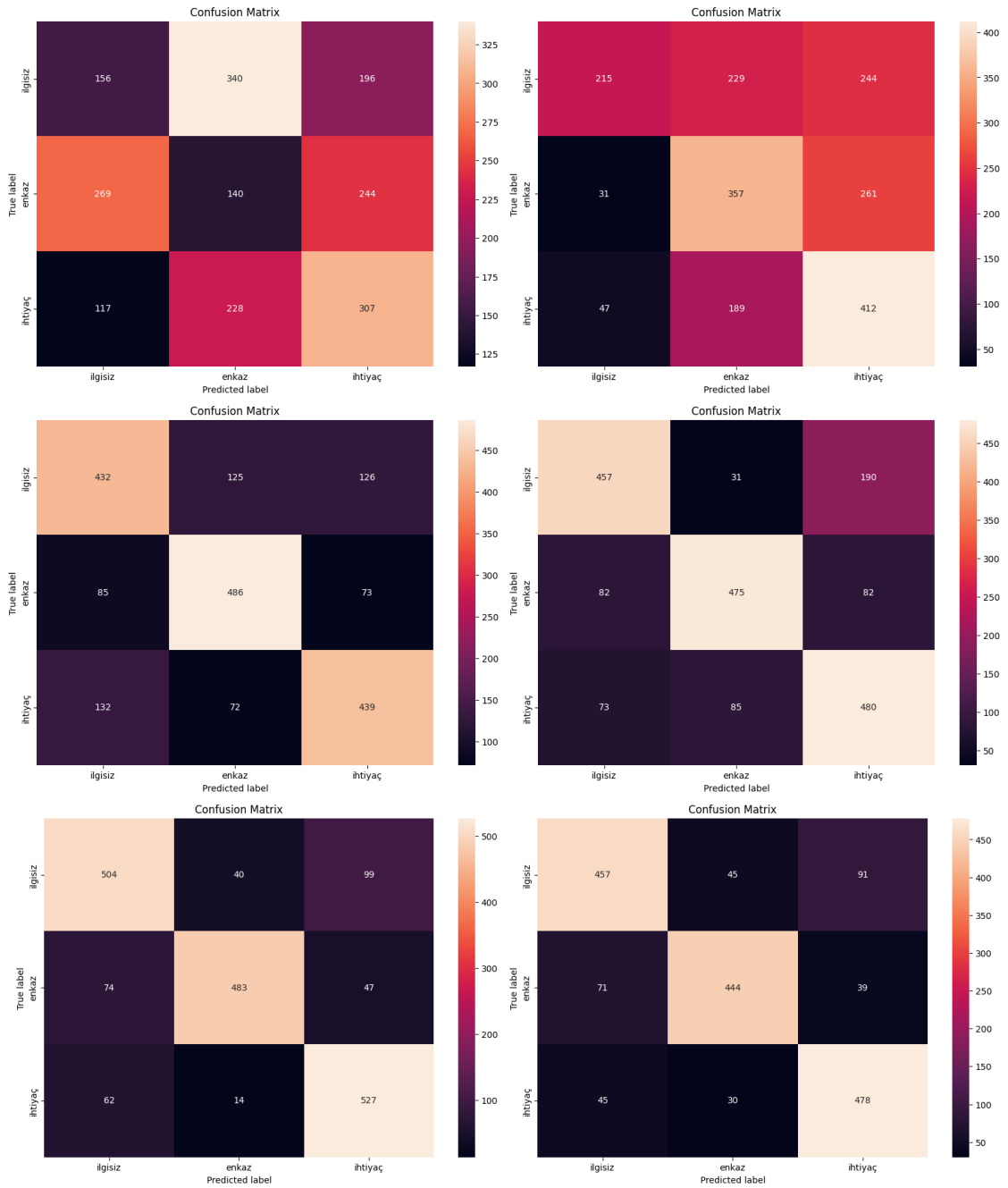
**Şekil 25.** BERT(ft) farklı veri seti ile eğitim sonrası birkaç örnek öğrenme yöntemi ile karmaşıklık matrisi (K=1,5,10,15,50,100)

Model CrisisNLP veri seti ile eğilirken amaç, eldeki az veriye rağmen benzer konuya ait başka veri seti ile ince ayar yapıldığında daha başarılı sonuçlar almaktır. Fakat gelen sonuçlara bakıldığında bu işlemin sonuçlarda büyük değişiklik sağlamadığı görülmüştür. Farklı bir konuya ait veri seti ile ince ayar yapıldıktan sonra 6 Şubat Depremi ile ilgili oluşturulan veri setinin sınıflandırma sonuçları ele alınarak benzer konulara ait veri seti kullanıp kullanmamanın etkisi incelenmiştir. Sonuçlar Tablo 12’te verilmiştir.

**Tablo 12.** BERT(Ft) Farklı Konuda Veri Seti ile Eğitim Sonrası Birkaç Örnek Öğrenme Sonuçları

|                  | K   | Doğruluk       | Veri Seti                     |
|------------------|-----|----------------|-------------------------------|
| <b>BERT (ft)</b> | 1   | %30,19         | Ürün Yorumları + 6 Şubat (ft) |
|                  | 5   | %49,57         | Ürün Yorumları + 6 Şubat (ft) |
|                  | 10  | %68,88         | Ürün Yorumları + 6 Şubat (ft) |
|                  | 15  | %72,22         | Ürün Yorumları + 6 Şubat (ft) |
|                  | 50  | %81,83         | Ürün Yorumları + 6 Şubat (ft) |
|                  | 100 | % <b>82,01</b> | Ürün Yorumları + 6 Şubat (ft) |

Bu sonuçlara ait karmaşıklık matrisleri Şekil 26’da verilmiştir.



**Şekil 26.** BERT(ft) Farklı konuda veri seti ile eğitim sonrası birkaç örnek öğrenme karmaşıklık matrisi (K=1,5,10,15,50,100)

CrisisNLP veri seti (2 etiketli) ile eğitim sonrası yapılan birkaç örnek öğrenme sonuçları ile ürün yorumlarından oluşan veri seti (3 etiketli) ile eğitim sonrası yapılan birkaç örnek öğrenme sonuçları incelenmiştir. Sonuç olarak 3 etiketli veri seti ile eğitim yapıldıktan sonra alınan sonuçlarda etiketlerin biraz daha iyi ayrıldığı gözlemlenmiştir. Özellikle ilgisiz ve ihtiyaç sınıfları arasındaki ayırım daha ince nüanslara dayanmaktadır. Bu kısımda 3 etiketli ürün yorumları veri seti ile eğitimin yapılması etiketlerin ayırımında faydalı olmuştur.

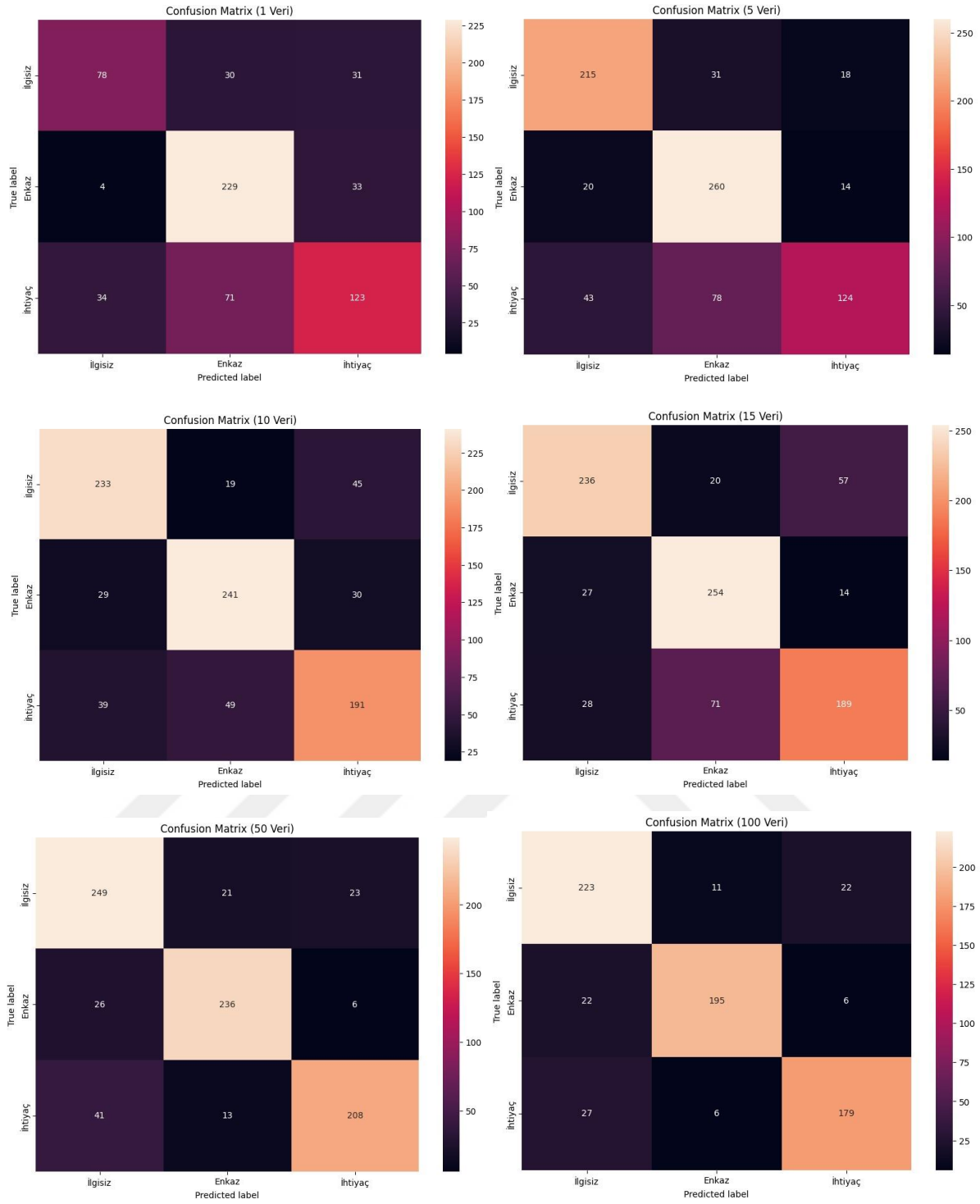
### TARS için sınıflandırma sonuçları

Çalışmanın bu kısmında TARS modeli için deneysel sonuçlar alınmıştır. Model kapsamında öncelikle oluşturulan 6 Şubat 2023 deprem veri setine paralel içeriği olan CrisisNLP veri seti kullanılmıştır. Bu veri seti ile kaynak görev oluşturulmuş ve eğitim yapılmıştır. Bu eğitim sonucunda elde edilen ağırlıklar kullanılarak 6 Şubat 2023 deprem veri seti ile birkaç örnek öğrenme işlemi yapılmış ve sonuçlar elde edilmiştir. Elde edilen sonuçlar Tablo 13'te gösterilmiştir.

**Tablo 13.** TARS Modeli Benzer Veri Seti ile Eğitim Sonrası Birkaç Örnek ile Öğrenme Sonuçları

|             | K   | Doğruluk | Veri Seti                |
|-------------|-----|----------|--------------------------|
| <b>TARS</b> | 1   | %28,28   | CrisisNLP + 6 Şubat (ft) |
|             | 5   | %51,29   | CrisisNLP + 6 Şubat (ft) |
|             | 10  | %58,49   | CrisisNLP + 6 Şubat (ft) |
|             | 15  | %63,78   | CrisisNLP + 6 Şubat (ft) |
|             | 50  | %75,31   | CrisisNLP + 6 Şubat (ft) |
|             | 100 | %81,10   | CrisisNLP + 6 Şubat (ft) |

Alınan sonuçlara bakıldığında sonuçların başarımların değerleri örnek sayısı ile doğru orantılı bir şekilde artmıştır. 6 Şubat Deprem veri setine paralel bir veri seti ile eğitim sonrasında birkaç örnek öğrenme işlemi yapılmasına rağmen 1 örnek sonrasında alınan sonuca bakıldığında etkisinin çok büyük olmadığı gözlemlenmiştir. Her örnek sayıları sonucunda alınan karmaşıklık matrisleri Şekil 27'de verilmiştir.



**Şekil 27.** TARS modeli benzer veri seti ile eğitim sonrası birkaç örnek ile öğrenme karmaşıklık matrisi sonuçları

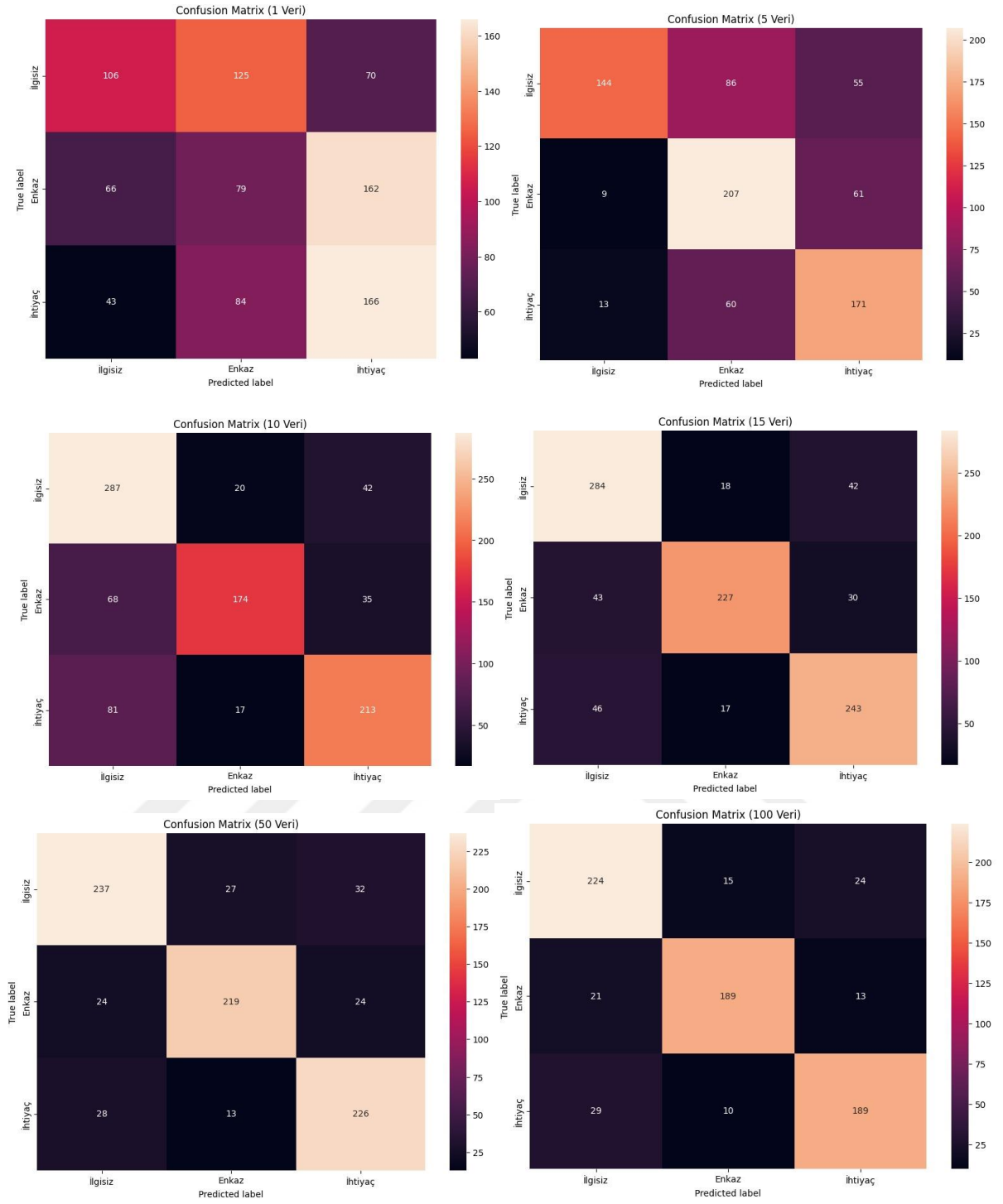
Karmaşıklık matrisleri incelendiğinde 1 örnek verildiği durumda enkaz sınıfına ait örnekler diğer sınıflara göre oldukça net bir şekilde ayırt edilmiştir. Buna rağmen ilgisiz ve ihtiyaç sınıflarında bir karışıklık söz konusudur. Bunun sebebi ilgisiz ve ihtiyaç sınıfı arasında ayırımın oldukça ince olmasından kaynaklanmaktadır. Örnek sayıları arttıkça bu ayırım netleşmiş ve sonuçlara yansımıştır.

Diğer bir adım olarak CrisisNLP veri seti gibi 6 Şubat Deprem veri setine paralel olmayan başka bir veri seti ile eğitilerek bunun ağırlıklarının alınmasının sonuçları incelenmiştir. Bu kısımda BERT(ft)'de olduğu gibi Türkçe ürün yorumlarından oluşan 3 etiketli veri seti kullanılmıştır. Örnek ve çevrim sayıları bir önceki adım ile aynı seçilmiştir. Sonuçlar Tablo 14'te verilmiştir.

**Tablo 14.** TARS Modelinde Farklı Konuda Veri Seti ile Eğitim Sonrası Birkaç Örnek Öğrenme Sonuçları

|             | K   | Doğruluk | Veri Seti                     |
|-------------|-----|----------|-------------------------------|
| <b>TARS</b> | 1   | %31,32   | Ürün Yorumları + 6 Şubat (ft) |
|             | 5   | %50,69   | Ürün Yorumları + 6 Şubat (ft) |
|             | 10  | %66,23   | Ürün Yorumları + 6 Şubat (ft) |
|             | 15  | %74,69   | Ürün Yorumları + 6 Şubat (ft) |
|             | 50  | %76,80   | Ürün Yorumları + 6 Şubat (ft) |
|             | 100 | %82,48   | Ürün Yorumları + 6 Şubat (ft) |

Alınan bu sonuçlar doğrultusunda 6 Şubat Deprem veri setine paralel konuları içeren veri setinin kullanılması ile konudan tamamen farklı Türkçe ürün yorumları veri setinin kullanılması arasındaki fark büyük değildir. Fakat burada 2 etiketli paralel konuya sahip veri setine göre az da olsa daha iyi sonuç veren 3 etiketli hedef konudan bağımsız veri seti dikkat çekmektedir. Bu sonuçlara göre etiket sayısının veri setinin içeriği kadar önemli olduğu gözlemlenmiştir. Bu deneye ait karmaşıklık matrisleri incelendiğinde ise 1 örnek verildiği durumda benzer konulu veri seti ile yapılan çalışmada enkaz sınıfının daha iyi tahmin edildiği görülmüştür. Farklı konuda veri seti ile eğitilmesi 3 etiketli olduğu için etiketlerin daha eşit dağılmasına neden olmuştur. Karmaşıklık matrisleri Şekil 28'de verilmiştir.



**Şekil 28.** TARS modeli farklı konulu veri seti ile eğitim sonrası birkaç örnek ile öğrenme karmaşıklık matrisi sonuçları

TARS modeli için çapraz doğrulama (cross validation) yapılmıştır. Çapraz doğrulama için 6 Şubat Deprem veri setinin tamamı 5 fold'a ayrılarak kullanılmıştır. Sonuçları Tablo 15'de verilmiştir.

**Tablo 15. TARS Çapraz Doğrulama Sonucu**

|             | Doğruluk | Veri Seti      | Yöntem                                             |
|-------------|----------|----------------|----------------------------------------------------|
| <b>TARS</b> | %84,95   | 6 Şubat Deprem | Çapraz doğrulama<br>(Cross Validation)<br>(5-fold) |

Bu tez çalışmasında yapılan çıkarımlar aşağıda listelenmiştir.

- Geleneksel yöntemlerin uygulanmasında 6 Şubat Deprem verisinin tamamı kullanıldığında alınan sonuçlar transfer öğrenme yöntemlerine oldukça yakın çıkmıştır. Fakat burada geleneksel yöntemlerin veri setini ezberlemeye oldukça yatkın olduğu unutulmamalıdır.
- Geleneksel yöntemlerde veri setinin tamamı kullanılmayıp birkaç örnek ile öğrenme kısmında yapıldığı alınan sonuçlarda düşüş yaşanmıştır. Buda geleneksel yöntemlerde veri sayısının önemini gözler önüne sermektedir.
- Geleneksel yöntemler için her şartta en iyi sonucu Lolistik Regresyon algoritması vermiştir.
- Tez kapsamında yapılan çalışmalarda 6 Şubat Deprem verisinin tamamen kullanıldığı uygulamalarda en iyi doğruluk oranını veren transfer öğrenme yöntemlerinden olan BERT olmuştur.
- Sıfır atış ile yapılan deneyde alınan sonuç başarılı demek için yeterli değildir.
- BERT (ft) modelinde en iyi sonuç farklı veri setleri ile ön eğitim yapılmadan, sadece 6 Şubat Deprem verisinden birkaç örnek öğrenme yöntemi ile alınmıştır.
- BERT(ft) modelinde benzer konulu veri seti ve farklı konu içeren veri etleri ile yapılan deneylerde sonuçlar çok yakın değerler olarak alınmıştır. Bu durum veri setinin içeriği kadar etiket sayılarının benzerliğinin de önemini göstermektedir.
- Birkaç örnek ile öğrenme modellerinin tamamında örnek sayılarının artması ile doğru orantılı olarak başarı sonuçları da artmaktadır. Buda örnek sayısı arttıkça modelin problemi ve etiketleri daha iyi tanıdığının bir göstergesidir.
- TARS modelinde yapılan birkaç örnek ile öğrenme sonuçlarına bakıldığında; benzer konulu 2 etiketli veri seti ile yapılan deneyde 1 örnek ile alınan sonuç, farklı konulu 3 etiketli veri seti ile yapılan deneyde 1 örnekte alınan sonuçtan daha düşük çıkmıştır. Buda etiket sayısının problemdeki etiket sayısına benzediğinde daha başarılı sonuçlar vermesinin bir göstergedir.
- Birkaç örnek öğrenme modellerinde örnek sayılarının artması her durumda başarıyı artırmıştır. Farklı veri setleri ile ince ayar yapıp ardından 6 Şubat deprem

verisinden örnekler verilerek yapılan deneylerde 1 örnekte farklılıklar olmasına rağmen örnek sayısı arttıkça sonuçlar birbirlerine oldukça benzemektedirler.



## SONUÇ VE ÖNERİLER

Bu çalışma, az etiketli veriyle deprem sonrası hızlı bir şekilde tweet sınıflandırması yapılabileceğini göstermiştir. Kullanılan tüm modeller incelendiğinde etiketli veri sayısının daha az olması durumları da dahil olmak üzere belirli bir seviyenin üzerinde başarılı sonuçlar elde edilmiştir. Bu özellikle deprem gibi doğal afet durumlarında, hızlı ve etkili bir şekilde toplumdan gelen geri dönüşleri analiz etmek için değerli bir araç olabilir.

Buna ek olarak, farklı veri setlerinin belirlenen hedef doğrultusundaki etkileri de incelenmiştir. Bu inceleme sonucunda ince ayar yapılan veri setinin konusu veya etiket sayılarının etkileri gözler önüne serilmiştir. Bu durumda daha farklı veri setleri ile de deneyler genişletilerek daha başarılı sonuçlar elde etmek için çalışılabilir.

Tezin hedefi olan veri kirliliğini azaltma ve hızlı aksiyon alabilme amacı göz önünde bulundurulduğunda, bu modellerin sınıflandırma yetenekleri, özellikle kriz anlarında büyük veri yığınlarından önemli bilgileri hızla çıkarma potansiyeli taşımaktadır. Bu hem hükümetlerin hem de ilgili kuruluşların, deprem sonrası duruma hızla müdahale etmelerine yardımcı olabilir.

Son olarak, bu çalışmanın sonuçları, dil modellemesi ve metin sınıflandırmasının, özellikle deprem gibi acil durumlar için toplumsal geri dönüşleri anlama konusunda ne kadar değerli olabileceğini göstermektedir. Öneri olarak, gelecekteki çalışmalarda, bu modellerin performansını artırmak için daha fazla etiketli veriyle eğitilmesi ve farklı metin ön işleme tekniklerinin uygulanması düşünülebilir. Bunun yanı sıra eğitim için farklı metriklerde dikkate alınabilir. Örneğin tweetin atıldığı hesabın doğruluğunun belirlenmesi ve eğitim yapılırken bunun dikkate alınarak yapılması daha hassas bir sınıflandırma yapılıp yapılmayacağı incelenebilir. Buna ek olarak zero-shot ile alınan sonucun geliştirilmesi için çalışılabilir ve farklı modeller oluşturulabilir. Yapılan sınıflandırma sonucunda oluşan sınıfları analiz ederek daha değerli hale getirilebilir. Örneğin enkaz altında kalan kişiler için yapılan sınıflandırmadan adresler tespit edilerek bunların bir haritada işaretlendiği çalışma yapılmıştı. Bunu geliştirerek ihtiyaç için toplanma alanları veya erzak dağıtım noktaları da bir harita üzerinde işaretlenebilir. Sadece deprem odaklı düşünmeyip modeller geliştirilerek ve hatta etiketler revize edilerek tüm doğal afet durumların başarılı sonuçlar verecek modeller geliştirilebilir. Geliştirilen bu modeller yapay zekâ uygulamalarında, robotikte veya resmi kuruluşlarda kullanılabilir.

## KAYNAKÇA

- A. C. Tantığ , "Metin Sınıflandırma", Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, c. 5, sayı. 2, ss. , Haz. 2016
- Akbik, Alan, *et al.* "FLAIR: An easy-to-use framework for state-of-the-art NLP." *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations)*. 2019.
- Alam, F., & Imran, M. (2020). CrisisDPS: Crisis Data Processing Services. In Proceedings of the 17th International Conference on Information Systems for Crisis Response and Management.
- Aldamen, Y.; Hacimic, E. Positive Determinism of Twitter Usage Development in Crisis Communication: Rescue and Relief Efforts after the 6 February 2023 Earthquake in Türkiye as a Case Study. *Soc. Sci.* **2023**, *12*, 436. <https://doi.org/10.3390/socsci12080436>
- American Red Cross (2010). "Social Media in Disasters and Emergencies". Erişim adresi: <http://i.dell.com/sites/content/shared-content/campaigns/en/Documents/Red-CrossSurvey-Social-Media-in-Disasters-Aug-2010.pdf>
- Argın, Y. (2023). Doğal Afetlerde Sosyal Medya Kullanımı: 2023 Kahramanmaraş Depremi Özelinde Twitter Örneği. *İnsanat Sanat Tasarım ve Mimarlık Araştırmaları Dergisi*, 3(1), 140-165.
- Avvenuti, M., Cresci, S., Marchetti, A., Meletti, C., & Tesconi, M. (2014). EARS (Earthquake Alert and Report System) a real time decision support system for earthquake crisis management. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 1749-1758.
- Barış Koçer, "Transfer Öğrenmede Yeni Yaklaşımlar", Selçuk Üniversitesi, Doktora Tezi, 2012.
- Bilgin, M. (2017). Gerçek veri setlerinde klasik makine öğrenmesi yöntemlerinin performans analizi. *Breast*, 2(9), 683.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Fleiss, Myunghee Cho Paik, Bruce Levin. 2003. Statistical Methods for Rates and Proportions. 3rd ed. John Wiley; Sons, Inc.
- Halder, Kishaloy, *et al.* "Task-aware representation of sentences for generic text classification." *Proceedings of the 28th International Conference on Computational Linguistics*. 2020.
- Houlsby, N., Giurigu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., ... & Gelly, S. (2019, May). Parameter-efficient transfer learning for NLP. In International Conference on Machine Learning (pp. 2790-2799). PMLR.
- [https://crisisnlp.qcri.org/crisis\\_datasets\\_benchmarks](https://crisisnlp.qcri.org/crisis_datasets_benchmarks)
- Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(3-4), 248-260.

- Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013, May). Practical extraction of disaster-relevant information from social media. In *Proceedings of the 22nd international conference on world wide web* (pp. 1021-1024).
- Kaggle deprem veri seti <https://www.kaggle.com/datasets/swaptr/turkey-earthquake-tweets>
- Kaggle ürün yorumları veri seti <https://www.kaggle.com/datasets/mujdatcabuk/eticaret-urun-yorumlari>
- Küçük, Doğan; Arıcı, Nursal. Doğal Dil İşlemede Derin Öğrenme Uygulamaları Üzerine Bir Literatür Çalışması. *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, 2018, 2.2: 76-86.
- Larochelle, H., Erhan, D., & Bengio, Y. (2008, July). Zero-data learning of new tasks. In *AAAI* (Vol. 1, No. 2, p. 3).
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach.
- Martinez-Gomez, P., Papachristoudis, G., Blauvelt, J., Rachlin, E., & Simhon, S. (2021). Enhancement and analysis of TARS few-shot learning model for product attribute extraction from unstructured texts.
- Meier, P. (2015). Digital humanitarians: how big data is changing the face of humanitarian response. CRC Press.
- Minaee, Shervin, *et al.* "Deep learning--based text classification: a comprehensive review." *ACM computing surveys (CSUR)* 54.3 (2021): 1-40.
- Muhammad Imran, Prasenjit Mitra, and Carlos Castillo. 2016. Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1638–1643, Portorož, Slovenia. European Language Resources Association (ELRA).
- Nguyen, D. T., Ofli, F., Imran, M., & Mitra, P. (2017). Damage assessment from social media imagery data during disasters. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.
- Nozza, Debora *et al.* "What the [MASK]? Making Sense of Language-Specific BERT Models." *ArXiv abs/2003.02912* (2020): n. pag.
- Ofli, F., Alam, F., & Imran, M. (2020). Analysis of social media data using multimodal deep learning for disaster response. *arXiv preprint arXiv:2004.11838*.
- Olteanu, A., Castillo, C., Diaz, F., & Vieweg, S. (2014). CrisisLex: A lexicon for collecting and filtering microblogged communications in crises. In *Proceedings of the 8th International Conference on Weblogs and Social Media*.
- Pan, S. J. and Yang, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345-1359, 2010.
- Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (Vol. 242, No. 1, pp. 29-48).
- Romain Meunier, Leila Moudjari, Farah Benamara, Véronique Moriceau, Alda Mari, *et al.*. Classification de tweets en situation d'urgence pour la gestion de crises. 18e Conférence en Recherche d'Information et Applications -- 16e Rencontres Jeunes Chercheurs en RI -- 30e Conférence sur le Traitement Automatique des Langues Naturelles -- 25e

- Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues, 2023, Paris, France. pp.204-216. {hal-04130181}
- S. Yazıcı and O. Zincir , "Kriz Yönetimi Ve Afetlerde Sosyal Medya Kullanımı", *İstanbul Üniversitesi Siyasal Bilgiler Fakültesi Dergisi*, vol. 0, no. 49, pp. 0, Jul. 2014
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. Proceedings of the 19th international conference on World wide web, 851-860.
- SALUR, Mehmet Umut; AYDIN, İlhan. The impact of preprocessing on classification performance in convolutional neural networks for Turkish text. In: *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*. IEEE, 2018. p. 1-4.
- Salvador Medina Maza, Evangelia Spiliopoulou, Eduard Hovy, and Alexander Hauptmann. 2020. Event-Related Bias Removal for Real-time Disaster Events. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3858–3868, Online. Association for Computational Linguistics.
- Sanh, V. D. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.
- Sui, D., Chen, Y., Mao, B., Qiu, D., Liu, K., & Zhao, J. (2020). Knowledge guided metric learning for few-shot text classification. arXiv preprint arXiv:2004.01907.
- Tien Nguyen, Dat, *et al.* "Rapid classification of crisis-related data on social networks using convolutional neural networks." *arXiv e-prints* (2016): arXiv-1608.
- TORAMAN, Cagri, *et al.* Tweets under the rubble: Detection of messages calling for help in earthquake disaster. *arXiv preprint arXiv:2302.13403*, 2023.
- U. Sahin, I. E. Kucukkaya, O. Ozcelik and C. Toraman, "Zero and Few-Shot Hate Speech Detection in Social Media Messages Related to Earthquake Disaster," 2023 31st Signal Processing and Communications Applications Conference (SIU), Istanbul, Turkiye, 2023, pp. 1-4, doi: 10.1109/SIU59756.2023.10224056.
- Wang, Y., Yao, Q., Kwok, J. T., & Ni, L. M. (2020). Generalizing from a Few Examples: A Survey on Few-shot Learning. *ACM Computing Surveys*, 53(3), Article No. 63. <https://doi.org/10.1145/3386252>
- Zhang, Y., & Vos, M. (2019). Social media monitoring: aims, methods, and challenges for international companies. *Corporate Communications: An International Journal*.
- ZHOU, Yangming, *et al.* A taxonomy of label ranking algorithms. *J. Comput.*, 2014, 9.3: 557-565.
- Ziletti, Angelo, *et al.* "Medical coding with biomedical transformer ensembles and zero/few-shot learning." *arXiv preprint arXiv:2206.02662* (2022).

## ÖZGEÇMİŞ

|                                      |                       |
|--------------------------------------|-----------------------|
| <b>Kişisel Bilgiler</b>              |                       |
| <b>Adı Soyadı:</b>                   | Dilara CANDAN KARADAŞ |
| <b>Doğum tarihi:</b>                 |                       |
| <b>Doğum Yeri:</b>                   |                       |
| <b>Uyruğu:</b>                       |                       |
| <b>Adres:</b>                        |                       |
| <b>Tel:</b>                          |                       |
| <b>E-mail:</b>                       |                       |
| <b>Eğitim</b>                        |                       |
| <b>Lise:</b>                         |                       |
| <b>Lisans:</b>                       |                       |
| <b>Yüksek lisans:</b>                |                       |
| <b>Doktora:</b>                      |                       |
| <b>Yabancı Dil Bilgisi</b>           |                       |
| <b>İngilizce:</b>                    | İyi                   |
| <b>Üye Olunan Mesleki Kuruluşlar</b> |                       |
|                                      |                       |
| <b>Tezden Üretilmiş Yayınlar</b>     |                       |
|                                      |                       |